

Propuesta para la enseñanza de la estadística inferencial informal en grado décimo

**Alejandra Hernández Patiño
Julián Alexander Robayo Suárez**

**Universidad Pedagógica Nacional
Facultad de Ciencia y Tecnología
Departamento de Matemáticas
Licenciatura en Matemáticas
Bogotá D.C. 2022**

Propuesta para la enseñanza de la estadística inferencial informal en grado décimo

Alejandra Hernández Patiño
C.C. 1024589776
Código: 2016240089

Julián Alexander Robayo Suárez
C.C. 1032477068
Código: 2017140067

Trabajo de grado presentado como requisito para optar por el título de Licenciado (a) en
Matemáticas

Director
Felipe Jorge Fernández Hernández
Magister en Estadística

Universidad Pedagógica Nacional
Facultad de Ciencia y Tecnología
Departamento de Matemáticas
Licenciatura en Matemáticas
Bogotá D.C. 2022

Dedicatoria

A la vida y a las personas que hicieron parte del proceso apoyando y aportando a que cada día seamos una versión mejorada de nosotros mismos.

-Alejandra y Julián.

Agradecimientos

Al profesor Felipe Fernández por su permanente disposición y dirección de este trabajo.

Al profesor Oscar Sosa y los estudiantes del proyecto por permitirnos el espacio, tiempo y participación para la aplicación de la secuenciación de las actividades y tareas propuestas.

Tabla de contenido

Dedicatoria.....	3
Agradecimientos.....	4
Tabla de contenido	5
Índice de tablas.....	7
Índice de figuras	8
Resumen.....	10
Introducción.....	11
Presentación.....	13
Descripción de la problemática o temática a abordar	13
Justificación	13
Objetivos	14
General.....	14
Específicos.....	14
Marco conceptual.....	16
Aspectos curriculares.....	16
Elementos disciplinares.....	17
Concepto de distribución.....	17
Población y muestra.....	18
Variación y noción de distribución.....	18
Centralidad.....	19
Dispersión.....	20
Ejemplo medidas de centralidad y dispersión	20
Distribuciones estadísticas.....	22
Distribuciones probabilísticas	23
Hipótesis estadísticas.....	25
Nivel de significación y significación estadística.....	26
Referentes de la educación estadística.....	26
¿Qué es la estadística inferencial informal?.....	26
Generalización más allá de los datos (G).....	27
Uso de los datos como evidencia (D).....	28
Articulación de la incertidumbre (LP).....	28
Razonamiento Informal Inferencial (RII).....	29
Comprensión, Interpretación y argumentación de la información estadística (CIA).....	30

Actitud crítica y cuestionadora (AC y C)	31
Dificultades y obstáculos asociados al aprendizaje y enseñanza de la IEI	31
Errores que se pueden cometer en el aprendizaje de la estadística inferencial informal .	32
Valoración del aprendizaje en la estadística inferencial informal	34
Marco metodológico	35
Contexto y estudiantes participantes	35
Prueba diagnóstica	36
Ruta del proyecto estadístico inferencial informal	37
Propuesta de actividades	38
Inferencia sobre proporciones	39
Inferencia de diferencia de proporciones	42
Inferencia sobre las medias poblacionales con una muestra de tamaño 100	43
Tipificación de las preguntas de la guía propuestas según los componentes del saber disciplinar y del saber matemático	45
Resultados	48
Análisis de los resultados de la prueba diagnóstica	48
Análisis de la aplicación de la secuencia didáctica	54
Guía uno	55
Guía dos	65
Guía tres	77
Conclusiones	85
Referencias	90
Anexos	92
1. Prueba diagnóstica	92
2. Guía uno	99
3. Guía dos	102
4. Guía tres	106

Índice de tablas

Tabla 1. Tipificación de las preguntas propuestas en la prueba diagnóstica.	37
Tabla 2. Componentes del saber pedagógico y disciplinar.....	47

Índice de figuras

Figura 1. Representación gráfica de una distribución de probabilidad	17
Figura 2. Tasa de mortalidad para COVID 19 por cada millón de habitantes en departamentos y distritos de Colombia al 21-02-2020.....	17
Figura 3. Representación gráfica de la distribución del ejemplo(a) y tabla resumen de las medidas de centralidad y dispersión (b).....	21
Figura 4. Representación gráfica de la regla empírica	24
Figura 5. Área bajo la distribución normal estándar $P(z \geq 1.58)$	25
Figura 6. Ruta del proyecto estadístico informal	37
Figura 7. Cálculo de proporciones muestrales en Excel para un tamaño de muestra $n=10$	56
Figura 8. Algunas hipótesis propuestas por los estudiantes para la muestra de tamaño $n=10$	57
Figura 9. Ejemplo del muestreo aleatorio simple por UPZ para un tamaño de muestra $n=10$	58
Figura 10. Argumentos sobre el muestreo inicial para un tamaño de muestra $n=10$. Resultados obtenidos con los estudiantes del proyecto.....	58
Figura 11. Gráfico estadístico y tabla resumen del análisis de datos consolidada con las muestras dadas por los estudiantes.....	59
Figura 12. Respuestas dadas por algunos estudiantes del proyecto a la pregunta 4 sobre la inferencia de proporciones para el conjunto de muestreos consolidados.....	60
Figura 13. Ejemplos de hipótesis para una muestra de tamaño $n=30$	62
Figura 14. Gráfico estadístico y tabla resumen de las proporciones muestrales consolidadas	63
Figura 15. Argumentos sobre la proporción muestral para un muestreo de tamaño $n=30$..	64
Figura 16. Hipótesis sobre la proporción muestral para un muestreo de tamaño $n=100$	65
Figura 17. Representación gráfica de la distribución y tabla resumen consolidadas para un muestreo de tamaño $n=100$	66
Figura 18. Validaciones sobre las hipótesis propuestas para una proporción muestral para un muestreo de tamaño $n=100$	67
Figura 19. Argumentos sobre los posibles valores máximos y mínimos del parámetro. Resultados de los estudiantes del proyecto	68

Figura 20. Comparación de las distribuciones para las proporciones de las 14 muestras de tamaños $n=10$, $n=30$ y $n=100$	69
Figura 21. Relaciones entre el tamaño del muestreo y las medidas de dispersión e intervalos de variación	70
Figura 22. Validaciones de las hipótesis estadísticas e inferencias sobre el parámetro estudiado.....	71
Figura 23. Hipótesis sobre diferencias de proporciones para muestras de tamaño $n=30$	72
Figura 24. Interpretación sobre la diferencia menor que cero.....	73
Figura 25. Diferencia de proporciones Patio Bonito – Américas para una muestra de tamaño $n=30$	74
Figura 26. Inferencia de diferencia de proporciones para una muestra de tamaño $n=100$..	74
Figura 27. Validación de hipótesis de diferencia de proporciones de tamaño $n=100$	75
Figura 28. Argumentación sobre el juicio final dado a las hipótesis planteadas para diferencia de proporciones	75
Figura 29. Argumentación de la contradicción encontrada por los estudiantes para el intervalo $(-0.1, 0.1)$	76
Figura 30. Ejemplo interpretación de la existencia del “0” en el intervalo $(-0.1, 0.1)$	76
Figura 31. Representaciones gráficas del muestreo conjunto y tabla de resumen para la inferencia de medias.....	77
Figura 32. Decisiones sobre la hipótesis sobre la media estudiada con base en el muestro de tamaño $n=100$	79
Figura 33. Conclusión sobre la hipótesis inicial de la edad promedio de las mujeres	80
Figura 34. Argumento sobre la menor probabilidad de edad	81
Figura 35. Argumentos sobre la obtención de una mejor inferencia.....	82
Figura 36. Ejemplo de respuestas para el ítem 7: mujeres de compras en Castilla.....	83
Figura 37. Progresión del desarrollo de los componentes de la IEI	83

Resumen

En el presente trabajo se propone el diseño de una secuencia de actividades que permitan potenciar el razonamiento inferencial informal en estudiantes de grado décimo, considerando particularmente los componentes del saber pedagógico específicos de la inferencia estadística informal: generalización más allá de los datos, uso de datos como evidencia y apropiación del lenguaje probabilístico.

Como resultados principales de la investigación se obtuvo: el impacto que tiene orientar este saber alrededor de una situación de interés de los estudiantes, el cual desemboca en la autonomía para la continuación del desarrollo de las guías orientadoras; el potencial de la interpretación de la representación gráfica de la distribución y los intervalos de variación como evidencias para establecer argumentos sólidos sobre la validación de las hipótesis; y, la comprensión de las diferencias entre los conceptos estimador y parámetro, descripción e inferencia, fruto de la interpretación del comportamiento de un conjunto de datos en la comparación de muestreos de diferente tamaño.

Introducción

En este trabajo de grado se presenta el diseño, desarrollo y análisis de un proyecto estadístico que se implementó con estudiantes de secundaria de grado décimo para promover la comprensión de elementos de la estadística inferencial de manera informal. La inferencia estadística informal (IEI) se puede definir como el razonamiento inductivo de los datos en general en el que se consideran las ideas generales para realizar conclusiones, utilizando argumentos que dan cuenta de la variabilidad, el análisis más allá de los datos y la consideración de los datos como evidencia.

A esta actividad estadística se le articula el razonamiento inferencial informal, al cual está relacionado con la destreza que tiene el estudiante para conectar ideas de distribución, centro y muestreo, destinado a validación de las hipótesis propuestas (Sánchez y Ruiz, 2017).

El desarrollo del trabajo se estructura en seis capítulos como se describe a continuación:

En la primera parte del trabajo se hace la presentación del proyecto en donde se describe la problemática o temática a abordar, la justificación y objetivos que dan cuenta a este trabajo.

En la segunda parte, se aborda el marco conceptual, conformado por los referentes curriculares, los referentes disciplinares y didácticos tenidos en cuenta para el diseño, aplicación y análisis de la prueba diagnóstica y la secuencia de actividades.

En la tercera parte se describe la aplicación de la prueba diagnóstica, tipificación de las preguntas propuestas según los componentes IEI, los componentes de la cultura estadística y el nivel de la lectura de datos y el análisis de los resultados de la prueba diagnóstica.

En la cuarta parte, se describe la secuencia de actividades implementadas alrededor de tres marcos disciplinares: inferencia sobre proporciones, inferencia sobre diferencia de proporciones e inferencia sobre medias poblacionales. En esta parte, también se muestra la

tipificación de las preguntas orientadoras, según los componentes pedagógicos y disciplinares tenidos en cuenta en el estudio.

Luego, se presenta el análisis de la aplicación de la secuencia de actividades (guías uno, dos y tres) realizando mediante una descripción breve de la interacción con los estudiantes, las evidencias de la aplicación y el análisis cognitivo en el cual se relacionan las respuestas dadas por los estudiantes, con la consulta bibliográfica.

Por último, se presentan las conclusiones alrededor del cumplimiento de los objetivos propuestos a lo largo de la investigación, así como las referencias y anexos.

Presentación

Descripción de la problemática o temática a abordar

Las investigaciones realizadas para la enseñanza y aprendizaje de la estadística suelen realizarse en mayor medida a nivel de educación básica que, a nivel educación secundaria (ver, por ejemplo, Batanero et al. (1993; citado en Fajardo, 2014)). Por ello, dichos autores afirman que los estudiantes en la educación superior presentan vacíos conceptuales y metodológicos en el aprendizaje de la estadística inferencial. En un sentido similar, Moreno y Vallecillos (2001; citado en Fajardo, 2014) sostienen que las dificultades que tienen los estudiantes en la educación superior están relacionadas con la complejidad del objeto y lenguaje matemático asociado, la falta de motivación y su baja capacidad para interpretar resultados de representaciones y resúmenes de datos.

En este sentido, se propuso el diseño, implementación y análisis de una propuesta didáctica enfocada hacia el aprendizaje de la estadística inferencial informal dirigida al grado décimo ya que, según los documentos curriculares, es en este grado en el cual los educandos tienen una experiencia cercana con la estadística inferencial (MEN, 2016).

En consonancia con lo anterior, por medio de esta propuesta de trabajo de grado se quiere responder al siguiente interrogante: ¿A través de qué tipo de actividades, tareas e investigaciones previas se puede promover la enseñanza de la estadística inferencial informal en el grado décimo?

Justificación

En los Estándares Básicos de Competencias para el desarrollo del pensamiento aleatorio y sistemas de datos en el grado décimo, se declara que los estudiantes deben ser capaces de diseñar experimentos aleatorios, proponer, justificar o refutar inferencias. Para ello deben hacer uso del razonamiento estadístico sobre datos reales y así validar las hipótesis planteadas, potenciando al estudiante a comprender y explicar el mundo que le rodea.

De esta manera, se fortalecen las destrezas relacionadas con la probabilidad, la estadística descriptiva, relaciones empíricas y lógica experimental implicadas en la estadística inferencial (v.g. pruebas de hipótesis y métodos de inferencia).

Además, según Batanero (2001) la relevancia de la enseñanza de la estadística mediante proyectos estadísticos permite contextualizar y motivar a los estudiantes a trabajar en un tema de interés, cuando se utilizan datos reales de un experimento. Además, se trabaja sobre la comprensión conceptual y de representaciones; la elección de un procedimiento según la situación; la resolución de problemas; la formulación y comunicación de las ideas matemáticas (hipótesis y conclusiones); el razonamiento matemático: el inductivo para reconocer patrones y plantear conjeturas, el deductivo para verificar conclusiones; por último, sobre la actitud con y hacia la estadística. Por tal motivo, su implementación ayudará a superar las dificultades asociadas al aprendizaje de la estadística inferencial informal.

Objetivos

General

Diseñar, proponer, implementar y analizar una secuencia de actividades y tareas que ejemplifiquen la enseñanza de la estadística inferencial informal en el grado décimo alrededor de una situación de interés de los estudiantes, para dar cuenta del desarrollo de las habilidades asociadas al pensamiento aleatorio y a los sistemas de datos enunciadas para este nivel formativo por los documentos curriculares.

Específicos

- ✓ Diseñar y proponer actividades y tareas que promuevan el razonamiento inferencial informal y el desarrollo de las habilidades relacionadas con los componentes específicos de la inferencia estadística informal.
- ✓ Implementar las actividades y tareas diseñadas en el grado décimo, y gestionar la actuación de los estudiantes alrededor de las dificultades, los obstáculos o los errores que tengan o puedan cometer.

- ✓ Analizar los resultados de la implementación de las actividades y tareas propuestas al contrastar los resultados con las intencionalidades enunciadas, a la luz de la literatura consultada.

Marco conceptual

Aspectos curriculares

Los Derechos Básicos de Aprendizaje (DBA) versión dos (MEN, 2016) para el desarrollo del pensamiento aleatorio y sistemas de datos enuncian que los estudiantes deben ser capaces de interpretar, valorar y analizar críticamente los resultados e inferencias presentadas en estudios estadísticos; comprender y explicar las medidas de tendencia central y de dispersión haciendo uso de los datos, tablas y gráficos para así obtener una mejor lectura de datos.

Las evidencias de aprendizaje que se tendrán en cuenta son:

- ✓ Identificar la población y las variables de estudio.
- ✓ Hacer inferencias sobre los parámetros basados en los estadígrafos calculados.
- ✓ Interpretar y comparar lo que representan algunas de las medidas de tendencia central y de dispersión en un conjunto de datos.
- ✓ Formular conclusiones sobre la distribución de un conjunto de datos, empleando más de una medida.

Por otro lado, los Estándares Básicos de Competencias en Matemáticas (EBCM) (MEN, 2006) para el desarrollo del pensamiento aleatorio y sistema de datos pronostican que los estudiantes son capaces de:

- ✓ Proponer inferencias a partir del estudio de muestras probabilísticas.
- ✓ Usar comprensiblemente algunas medidas de centralización, localización, dispersión y correlación (percentiles, cuartiles, centralidad, distancia, rango, varianza, covarianza y normalidad).
- ✓ Interpretar nociones básicas relacionadas con el manejo de información como población, muestra, variable aleatoria, distribución de frecuencias, parámetros y estadígrafos.
- ✓ Justificar o refutar inferencias basadas en razonamientos estadísticos a partir de resultados de estudio publicados en los medios o diseñados en el ámbito escolar.

Elementos disciplinares

Concepto de distribución

En matemáticas, en particular en el área de la probabilidad, se define una distribución como la función que representa las probabilidades que identifican y describen el comportamiento variable de un fenómeno de carácter aleatorio. Por ejemplo, la probabilidad de la distribución de carga total entre a y b , determinada por la integral de la función de densidad de a hasta b , se representa como el área bajo la curva en esos intervalos (v.g. Figura 1). Por otra parte, en la disciplina estadística, una distribución representa a una o varias variables de interés, en un lenguaje matemático que describe el comportamiento de un fenómeno social, económico y/o científico. En este sentido, la distribución puede ser utilizada en diversidad de ciencias y disciplinas. Por ejemplo, al realizar un estudio epidemiológico como el del COVID 19 (Figura 2), es pertinente conocer la distribución de este fenómeno para diseñar protocolos de prevención y de tratamiento de la enfermedad estudiada.

Figura 1

Representación gráfica de una distribución de probabilidad. Imagen recuperada de Anderson et al. (2008), mayo 2022

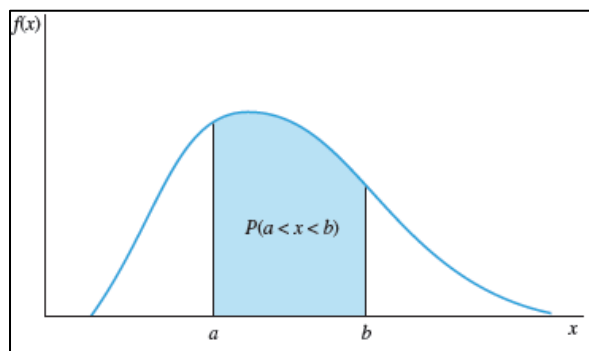
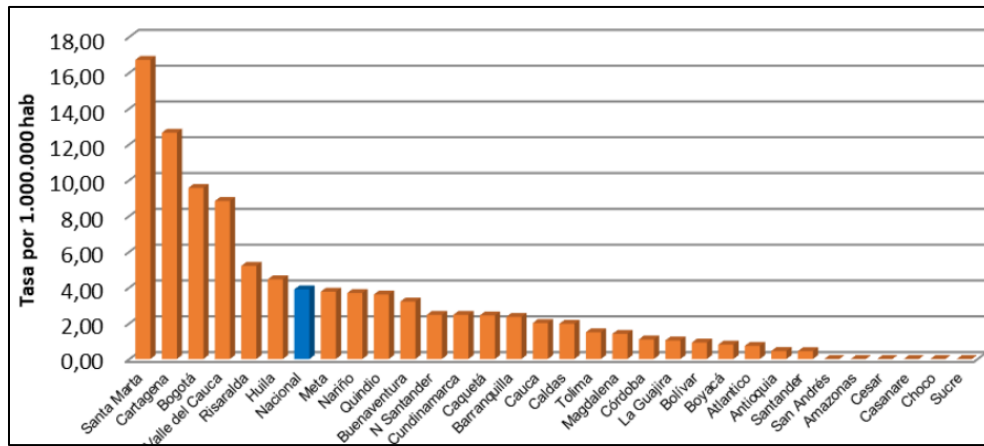


Figura 2.

Tasa de mortalidad para COVID 19 por cada millón de habitantes en departamentos y distritos de Colombia al 21-02-2020. Imagen recuperada de <https://www.ins.gov.co/Noticias/Paginas/Coronavirus.aspx>, mayo 2022, mayo 2022



Población y muestra

La población hace referencia al conjunto de todos los elementos del estudio determinado, mientras que la muestra es un subconjunto de la población. Cuando la recolección de los datos se hace para toda la población, se está realizando un censo, mientras que, si se recolectan los datos de una muestra, se realiza una encuesta muestral o muestreo. Debido al tiempo, costo u otros factores, en algunas situaciones es preferible realizar estimaciones sobre la población a partir de una muestra, probando así las hipótesis sobre la(s) variable(s) de interés de una población (Anderson et al., 2008). Por ejemplo, para las pruebas de sangre se extrae una muestra del individuo, esta se analiza y el profesional de la salud puede así interpretar los resultados obtenidos de esa muestra para caracterizar aspectos del cuerpo del sujeto analizado. Otro ejemplo, podría darse al realizar un estudio territorial en donde Colombia sería la población y Cundinamarca, una muestra.

Variación y noción de distribución

En muchas de las formas de razonamiento estadístico sobre la variabilidad aparece la noción de distribución; sobre esta subyacen grandes ideas de la estadística como el muestreo y el análisis de varianza (Bakker y Gravemeijer, 2004). Señalan Reading y Shaughnessy (2004) que “sin variación no hay distribución”. En reacción a esta afirmación, Wild (2006) recalca que para comprender la noción de distribución se debe comenzar por tener en cuenta la universalidad de la variabilidad. De hecho, para Wild (2006), la noción de distribución en su nivel más elemental es “el patrón de variación en una variable; esta noción es el fundamento del trabajo estadístico que investiga desentraña y modela patrones de variación, con el fin de aprender de ellos y hacerlos predecibles”. En este sentido, Wild (2006) establece la

distribución como un lente para mirar la variabilidad. Por lo tanto, para este autor hay una relación muy estrecha entre la variabilidad de una colección de datos que conforman una variable y la correspondiente distribución.

Hay varias nociones que se pueden asociar a una distribución. Por ejemplo, en Andrade, Fernández y Méndez (2021) mencionan las ideas de centralidad, dispersión, densidad y forma. En esta sección se mencionarán algunas de ellas.

Centralidad

La tendencia central o centralidad se refiere a la información que aportan diferentes indicadores cuantitativos que tienden a ubicarse en el centro de la distribución y que permiten resumir dicho atributo del conjunto de datos. Debido al mayor tratamiento escolar y al propósito que se persigue en el presente trabajo, se reconocerán la media, la mediana y la moda de los datos.

La media de un conjunto de datos que se suele representar con el símbolo “ \bar{x} ”, se define como la suma aritmética de los datos dividida por el total de datos. Es decir, dadas las observaciones x_1, x_2, \dots, x_n , la media o promedio aritmético se define como $\bar{x} = \sum_{i=1}^n \frac{x_i}{n}$.

La mediana de un conjunto de datos, que se denotará como Me busca identificar un valor intermedio de los datos al ordenarlos de manera creciente. Es decir, dadas las observaciones x_1, x_2, \dots, x_n , donde $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ son los mismos datos, pero ordenados de manera creciente, la mediana se puede definir como

$$Me = \begin{cases} x_{(\frac{n+1}{2})} & \text{si } n \text{ es impar,} \\ \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}) & \text{si } n \text{ es par.} \end{cases}$$

Por último, la moda, que se denotará como Mo , es un indicador que busca identificar el valor que se presenta con mayor frecuencia en la distribución de los datos (Anderson, Sweeney & Williams, 2008).

Dispersión

La dispersión o variación es un atributo de la distribución que hace referencia a la variabilidad de los datos, a las diferencias que existen entre ellos y, por tanto, a la mayor o menor representatividad de los promedios. Los indicadores bajo el cual se definen medidas de dispersión buscan medir el grado de homogeneidad o de heterogeneidad de los datos. Para efectos del presente trabajo, se consideran los siguientes indicadores de dispersión: rango, varianza, desviación estándar y rango intercuartílico del conjunto de datos.

El rango, es la medida más sencilla o elemental para medir dispersión, se define como el resultado de la diferencia entre el valor máximo y el valor mínimo del conjunto de y se puede denotar de manera breve como " $X_{máx} - X_{mín}$ ".

La varianza muestral de un conjunto de datos, usualmente denotada como S^2 , es la suma promediada de las desviaciones al cuadrado de las observaciones (x_i) respecto al promedio aritmético. y se expresa mediante la fórmula $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$. Por otra parte, la desviación estándar, se calcula como el valor absoluto de la raíz de la varianza. Aunque ambas, la varianza y la desviación estándar miden la variación de los valores mayores y menores respecto a la media aritmética, para propósitos de interpretación es preferible el uso de la desviación estándar, ya que esta se expresa en las mismas unidades de medida que la variable estudiada. Por otra parte, es pertinente señalar que, si los valores de la muestra son los mismos, entonces cualquier medida de dispersión daría cero, es decir, no hay variación entre los datos.

Ejemplo medidas de centralidad y dispersión

Una multinacional de tecnología quiere realizar un informe sobre la cantidad de celulares que se venden en las sucursales de Colombia en el día de las madres. Para ello, toma una muestra de 40 sucursales las cuales reportaron las siguientes unidades vendidas:

30	30	35	43	37	37	43	42	32	32	28	45
33	30	26	39	37	41	26	27	34	25	25	37
37	40	34	41	26	43	32	42	36	29	33	39
39	29	26	37								

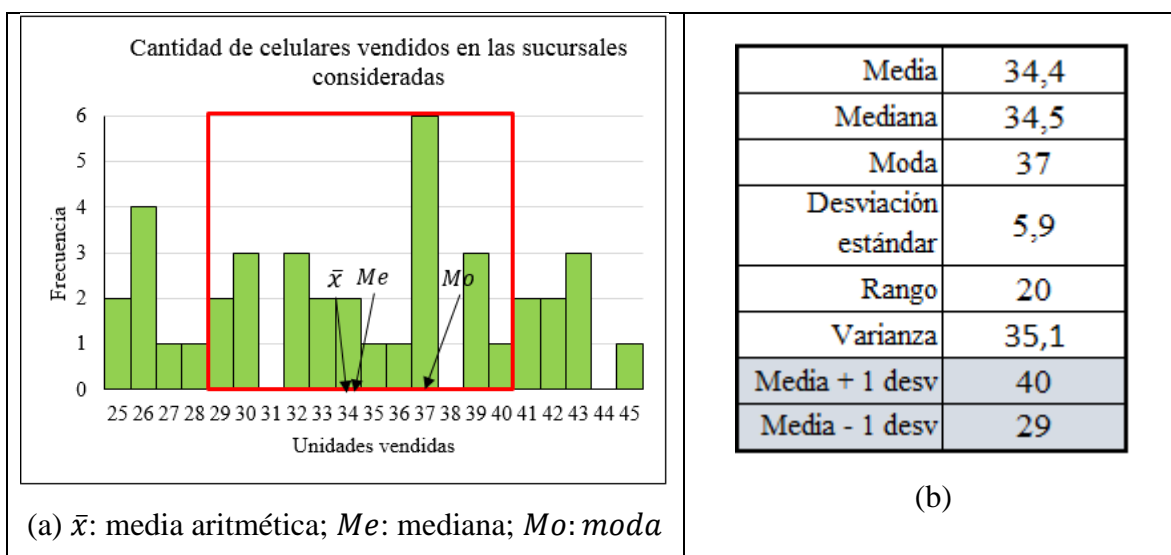
Siguiente, realiza la caracterización tomando en cuenta la media aritmética, la mediana, la moda, la desviación estándar y el rango de los datos recibidos. La distribución de los datos y el análisis estadístico se ilustran en la Figura 3.

La media de los datos es 34,4 pero como se venden unidades enteras, la empresa reporta que el promedio de celulares vendidos es de 35 unidades; la mediana de los datos es 34,5, ya que como hay 40 datos, se promedian los datos los cuales al organizarse están en la posición 20 y 21, es decir, $Me = \frac{34+35}{2} = 34,5$. En conclusión, más del 50% de las sucursales venden una cantidad mayor o igual a 35 celulares; luego, hallan la moda y reportan que las tiendas lo que más suelen vender es 37 celulares en el día de las madres.

A continuación, hallan el rango y la desviación estándar, siendo estos valores 20 y 5,9 respectivamente. En este orden de ideas, reportan que hay gran dispersión en la cantidad de celulares que se venden en las sucursales en Colombia consideradas como muestra el día de las madres y que, por norma empírica concluyen que alrededor del 60% de las sucursales suelen vender entre 29 y 40 unidades en este día.

Figura 3.

Representación gráfica de la distribución del ejemplo(a) y tabla resumen de las medidas de centralidad y dispersión (b)



Distribuciones estadísticas

En las distribuciones de tipo estadístico se reconoce que las variables de estudio pueden hacer referencia a una población, una muestra, o un muestreo. La de una población usualmente hace referencia a la distribución de una característica o variable asociada al total de miembros de un grupo o conjunto de interés; la de una muestra, hace referencia a la distribución de una característica o variable de una parte seleccionada de la población para analizar; y finalmente, la de muestreo, es la abstracción teórica, que resulta de la acción de elección dentro de un marco o lista de la población objetivo, de todas las muestra posibles de un tamaño dado a las que se les calcula una medida de resumen estadístico tal como un promedio, una proporción o una desviación.

Según el tipo de muestras utilizadas, las muestras pueden ser probabilísticas o no probabilísticas. En el muestreo no probabilístico, no se conoce la probabilidad de selección de los elementos seleccionados, son económicas y rápidas de obtener. No obstante, debido al sesgo y a la falta de generalización de las muestras seleccionadas, su uso debe limitarse a grupos pequeños de interés o a estudios que no requieran un nivel alto de rigurosidad. Entre las muestras no probabilísticas se distinguen las muestras de juicio, la cual consiste en las opiniones de los expertos en el tema de estudio y el muestreo de conveniencia, dónde los elementos de la muestra se consiguen de manera fácil y económica.

Las muestras probabilísticas se obtienen por medio del muestreo probabilístico, es decir, donde se conoce de antemano la probabilidad de elección de sus elementos respecto del marco poblacional de referencia. Entre otros métodos, la muestra se puede seleccionar por simple aleatoriedad, de manera sistemática, por estratificación o mediante conglomerados.

En el muestreo aleatorio simple, todos los elementos del espacio muestral tienen la misma probabilidad de ser elegidos. Además, este muestro puede ser con o sin reemplazo. En la muestra sistemática, los elementos del espacio muestral o marco se dividen en n grupos de k elementos, $k \in \mathbb{Z}$; el primer elemento se selecciona entre los primeros k elementos del espacio muestral (en el primer grupo), y los demás, serían k -ésimo elemento de los demás grupos. El muestro por estratificación consiste en dividir los elementos del espacio muestral

en subpoblaciones separadas con características distinguibles; de cada estrato se elige una muestra aleatoria simple y se combinan los posibles resultados distintos. El muestreo de conglomerados se suele realizar en el estudio de costo – eficacia sobre una región geográfica; en ella, se divide el marco en conglomerados (ej.: países, barrios, cuadras) y se seleccionan muestras aleatorias de cada subconjunto obtenido (Levine, Krehbiel & Berenson, 2006).

Distribuciones probabilísticas

Respecto al tipo de distribución de las muestras, según la naturaleza de la variable aleatoria de estudio, la distribución puede ser discreta o continua. La distribución de probabilidad para una variable aleatoria x se representa mediante una fórmula, tabla o gráfica en la que asocian los valores de la variable aleatoria con su probabilidad $p(x)$.

Por un lado, para que una distribución de probabilidad sea discreta, esta debe cumplir i) tomar valores entre 0 y 1 y ii) $\sum p(x) = 1$. Como ejemplo de distribución de probabilidad discreta se tiene la distribución binomial, ideal para representar la distribución de una prueba de Bernoulli, en la cual se consideran como resultados del experimento los éxitos o fracasos. Esta distribución está representada mediante la expresión $P(X = x) = B(x, n, p) = \binom{n}{x} p^x q^{n-x}$, donde n son las pruebas independientes, p la probabilidad de éxito, $q = 1 - p$, la probabilidad de fracaso, x la cantidad de éxitos a estudiar y n el total de ensayos. Además, la media o valor esperado y la varianza de la distribución binomial son $\mu = np$ y $\sigma^2 = npq$ respectivamente (Mendenhall, Beaver & Beaver, 2006).

Por otro lado, la distribución de probabilidad continua se caracteriza por ser descrita mediante una función de densidad de probabilidad de la variable de estudio de tal forma que $f(x)$; $P(a < x < b) = \int_a^b f(x)$, interpretada como el área bajo la curva entre el intervalo descrito (v.g. Figura 1). A diferencia de las variables aleatorias, $P(x = a) = 0$ porque se estaría considerando el área debajo de un punto. Lo anterior implica que $P(x \geq a) = P(x > a)$; $P(x \leq a) = P(x < a)$.

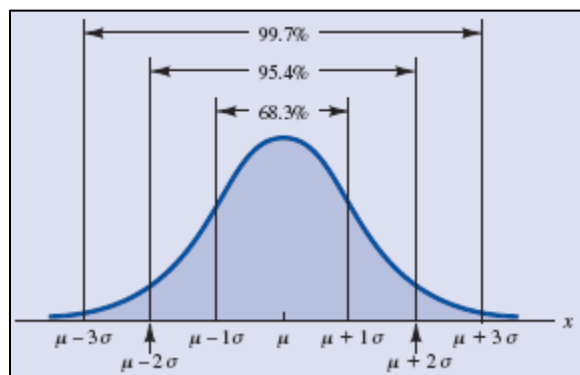
Si los valores de probabilidad están distribuidos uniformemente, la variable aleatoria de estudio se conocerá como variable aleatoria uniforme. En consecuencia, la función de

densidad se podría representar mediante una figura plana. Otro tipo de variable que suele presentarse en fenómenos sociales, económicos o de las ciencias naturales es el que se asocia a la distribución normal.

De hecho, en la inferencia estadística se utiliza con frecuencia la distribución normal de probabilidad cuya función de densidad es $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $x \in \mathbb{R}$, $\sigma > 0$; μ representa la media poblacional y σ , su desviación estándar. Esta función es simétrica respecto a la media μ , quién representa el centro de la distribución. Mientras que la forma de la distribución se describe mediante el parámetro σ , ya que si σ es grande, la altura de la curva se verá reducida y la dispersión aumentará; caso contrario, se obtendrá una curva con mayor altura y menor dispersión. Además, los estimadores representados mediante esta distribución siguen la regla empírica en la cual, aproximadamente el 95% de todas las estimaciones puntuales estarán ubicadas dentro del intervalo $\mu \pm 2\sigma$ (ver Figura 4).

Figura 4.

Representación gráfica de la regla empírica. Imagen recuperada de Anderson et al. (2008), mayo 2022

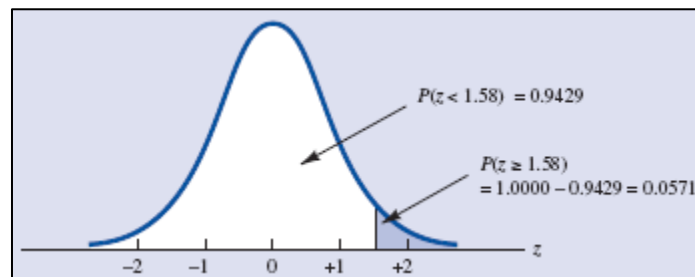


Debido a que no es sencillo realizar la integral para hallar los valores de probabilidad de la variable aleatoria normal y que se pueden obtener infinitas distribuciones, se han propuesto tablas estandarizadas para representar estas áreas bajo la curva. La estandarización de la variable normal estandarizada se representa por la variable z , siendo $z = \frac{x-\mu}{\sigma}$; $x = \mu + z\sigma$

(v.g. Figura 5). La distribución de esta variable se conoce como distribución normal estandarizada, para la cual $\mu = 0$ y $\sigma = 1$; $z < 0$ si $x < \mu$; $z > 0$ si $x > \mu$; $z = 0$ si $x = \mu$ (Mendenhall et al., 2006).

Figura 5.

Área bajo la distribución normal estándar $P(z \geq 1.58)$. Imagen recuperada de Anderson et al. (2008), mayo 2022



Hipótesis estadísticas

Una prueba estadística de hipótesis está formada por los siguientes elementos partes: la hipótesis nula, la hipótesis alternativa, el estadístico de prueba, el valor p del resultado del estadístico de prueba, la región de rechazo y la conclusión.

En general, al realizarse una hipótesis se inicia con una teoría, afirmación o una suposición tentativa acerca de un parámetro poblacional. A la suposición tentativa se le llama hipótesis nula y se denota como H_0 . Luego se define otra hipótesis llamada hipótesis alternativa, que dice todo lo contrario de lo que se establece en la hipótesis nula. La hipótesis alternativa se denota como H_a o H_1 .

Es necesario aclarar que nunca se sabe si la H_0 es verdadera o falsa. Sin embargo, en una prueba de hipótesis el investigador estadístico empieza por suponer que H_0 es verdadera, y luego, al utilizar los datos muestrales debe decidir si la evidencia está a favor de H_0 o H_a y se pueden obtener dos conclusiones:

- (1) Se rechaza H_0 y se concluye que H_a es verdadera, cuando los datos están en contravía de lo que se supuso.
- (2) No se rechaza H_0 y se asume que es verdadera, cuando los datos están en consonancia con lo que se supuso.

Nivel de significación y significación estadística

El nivel de significancia o valor p de una prueba estadística se refiere al valor más pequeño de un alfa (α) para el cual H_0 se puede rechazar. Se debe tener en cuenta que existe un riesgo real de cometer un error tipo uno, si H_0 es rechazada con base en el valor observado del estadístico de prueba. El valor p mide la fuerza de la evidencia contra H_0 .

Si el valor p es menor o igual a un nivel de significancia α asignado previamente, entonces H_0 puede ser rechazada y se puede informar que los resultados son estadísticamente significativos al nivel de α (Anderson et al., 2008).

De lo anterior se puede inferir que el investigador debe tener en cuenta ¿Qué tan pequeño debe ser el valor p antes que decía H_0 ? Y

- (1) Si el valor p es menor a 0.01, H_0 se rechaza. Los resultados son altamente significativos.
- (2) Si el valor p está entre 0.01 y 0.05, H_0 se rechaza. Los resultados son estadísticamente significativos.
- (3) Si el valor p está entre 0.05 y 0.10, por lo general no se rechaza. Los resultados son tendentes hacia significancia estadística.
- (4) Si el valor p es mayor a 0.10, H_0 no es rechazada. Los resultados no son estadísticamente significativos.

Referentes de la educación estadística

¿Qué es la estadística inferencial informal?

Según Ruiz, Batanero y Arteaga (2011; citado por Sánchez y Ruiz, 2017), la inferencia estadística (IE) representa la actividad principal en el análisis de datos. Por lo tanto, para dicha labor es importante utilizar e integrar conocimientos previos formales e informales que permiten interpretar los datos desde tal perspectiva. Mientras que los conocimientos formales se refieren a conceptos básicos como la noción de distribución y la aplicación e interpretación de medidas de resumen estadístico de centralidad y dispersión, entre otros, los informales se relacionan con aspectos como el conocimiento del contexto de los datos y de las muestras involucradas en una inferencia.

Para Makar & Rubin (2009) una inferencia estadística es una generalización de los datos disponibles, expresada en un lenguaje probabilístico, evidenciada, razonada y extendida más allá de los datos. En este sentido, la inferencia permite potenciar en los estudiantes la articulación de una predicción, desde las observaciones, la organización y el uso de datos, el trabajo de la variabilidad y la relación de los datos obtenidos con su contexto.

Ahora bien, la inferencia estadística informal (IEI), como lo sugieren Zieffler, Garfield, Delmas & Reading (2008), es la descripción de la manera en que los estudiantes hacen uso del conocimiento estadístico informal para entender los fenómenos que lo rodean, con base en muestras observadas. En este sentido, Makar & Rubin (2009), la contemplan como un proceso para aprender estadística el cual incluye el proceso de generalización o predicción, el uso de los datos y el lenguaje probabilístico. En decir que la IEI es un tipo de razonamiento que considera ideas para realizar conclusiones, utilizando argumentos que dan cuenta de la variabilidad y la incertidumbre, con base en un análisis que va más allá de los datos y que considera los datos como evidencia.

Autores como Sánchez y Ruiz (2017) aseveran que, en el proceso para hacer inferencias basadas en datos, se distinguen tres componentes claves: 1) la generalización o predicción al observar más allá de los datos, 2) el uso de los datos como evidencia, y 3) el empleo de un lenguaje probabilístico al describir generalizaciones a partir de las conclusiones propuestas. En lo que sigue, se esbozarán las características principales de tales componentes.

Generalización más allá de los datos (G)

Las generalizaciones estadísticas sobre las características de una población son abstracciones que se realizan partiendo de una muestra. Esta actividad implica hacer uso de los datos que se tiene a la mano (muestras) para poder hacer inferencias que vayan más allá de los datos, es decir, caracterizar poblaciones existentes o inexistentes, en términos, por ejemplo, de medidas de centralidad, proporción o dispersión. Estas generalizaciones pueden utilizarse para evaluar las hipótesis; es decir, el uso de métodos de inferencia estadística y el análisis realizado sobre la obtención de los datos se constituyen en herramientas para valorar la veracidad de tales hipótesis (Makar & Rubin, 2009).

Algunos autores (ver, por ejemplo, Batanero 2001) sostienen que en la educación estadística la IEI se puede potenciar mediante software estadísticos y actividades manipulables que le

posibiliten a los estudiantes configurar el tamaño de muestras a placer. De esta manera, concretan inferencias sobre la población soportándose de imágenes y trazos observables en el comportamiento de los datos.

Uso de los datos como evidencia (D)

Usar los datos como evidencia, conlleva la articulación de argumentos y justificaciones para poder hacer juicios, afirmaciones o predicciones sobre las poblaciones sustentadas con base en las muestras disponibles. Las argumentaciones y justificaciones pueden ser observacionales, numéricas, descriptivas o no registradas. La evidencia es fundamental para el desarrollo de la comprensión en los estudiantes para poder justificar las afirmaciones y las inferencias.

Desde las posturas actuales para la educación estadística se sostiene que los estudiantes aprenden más sobre los datos cuando se les motiva el asumir una actitud crítica sobre los mismos, que solo enfatizando en la realización de ejercicios aritméticos. No obstante, los estudiantes no están acostumbrados a conectar lo que conocen de su vida con lo que aprenden en el colegio. Esto se debe a que los estudiantes consideran los datos y el contexto de los datos por separado. Una de las sugerencias para atender a este problema es la de recurrir a simulaciones que permitan a los estudiantes razonar y confrontar sus ideas acerca de los datos y el contexto bajo el cual fueron tomados, en diferentes niveles de lectura.

Esa estrategia educativa pretende promover la utilización de los datos como evidencia de la generalización, según el nivel de incertidumbre que se entrevea en los datos. Así, comentando lo que dicen Sánchez y Ruiz (2017) sobre ese componente, se puede decir que los estudiantes van adquiriendo confianza al plantear sus conclusiones en la medida en que trabajen comparando una muestra con una distribución de muestreo.

Articulación de la incertidumbre (LP)

El uso de lenguaje probabilístico se evidencia en la expresión de términos o frases que modulan las afirmaciones y posibilitan hablar acerca de la incertidumbre presente en las conclusiones. Como en la inferencia estadística, las conclusiones que se realizan sobre la población se obtienen desde una muestra, es natural que éstas presenten cierto nivel de incertidumbre, porque la característica observada puede no aplicarse para todos los miembros

de dicha población. En consecuencia, es necesario nutrir el uso del lenguaje probabilístico con experiencias de incertidumbre aplicadas a hipótesis generadas, según el nivel de escolaridad de los estudiantes.

En altos niveles de escolarización, los estudiantes tienen una experiencia más cercana tratando la incertidumbre con niveles de confianza y márgenes de error en comparación al nivel primario, en donde las probabilidades de los niveles de confianza y la precisión a la que alude un margen de error, no se pueden expresar formalmente. En cualquier caso, la idea es que las secuencias de enseñanza que se propongan permitan evidenciar progreso en la adquisición de lenguaje probabilístico, y en la articulación de la incertidumbre con base en la evidencia que aporten los datos.

Razonamiento Informal Inferencial (RII)

Para Pfankuch (2006, citado en Sánchez y Ruiz, 2017), el razonamiento inferencial informal está comprendido por los procesos de pensamiento que se realizan para establecer hipótesis sobre la distribución, el muestreo y el centro. Tal razonamiento depende de las experiencias con el contexto de los datos y los datos en sí, en los que se utilicen conceptos estadísticos y herramientas digitales que posibiliten la representación gráfica y numérica que arrojen los datos como evidencia de las generalizaciones realizadas.

El RII se puede incitar o promover cuando se proponen situaciones que conlleven el razonar sobre un problema que pueda ser modelado para tratarlo con conceptos estadísticos, ya que implica utilizar los datos para realizar conclusiones contextuales. De este modo, se promueve que se generen procesos de integración del conocimiento estadístico y de lo que se conoce sobre el mundo natural.

Asimismo, Voss, Pekins & Segal (1991; citado en García y Sánchez, 2014) consideran que el RII se promueve cuando se le propone al estudiante situaciones no deductivas en contextos de la vida cotidiana. En este sentido, el RII tiene que ver con argumentaciones que se pueden evaluar en términos de la solidez de los siguientes aspectos:

- ✓ Valoración de si las razones que apoyan el argumento son verdaderas o aceptables.
- ✓ Consideración de la medida en que los argumentos apoyan la conclusión a la que llega el individuo.

- ✓ Revisión de la medida en que el individuo se defiende cuando se le presentan argumentos en contra, es decir, las razones que expresa para continuar apoyando su postura.

Como resultado de este razonamiento, se obtienen conclusiones lógicas a partir de los datos, soportadas a través de una argumentación valorable y basada en el análisis de datos (Garfield & Ben – Zvi, 2011; citado por García, 2014).

En esto sentido se podría anunciar que el RII gira alrededor del contexto de la situación problema. Según Dewey (1910; citado por Makar et al. 2009), en el contexto de una situación problema se pueden considerar tres roles en el acto de hacer inferencias. Primero, lleva a dudar o tener pensamientos de incertidumbre sobre el contexto del problema y, en consecuencia, la necesidad de investigarlo. Segundo, el contexto es necesario para darle sentido a la inferencia ya que pone en juego las normas y creencias de la persona sobre las hipótesis planteadas. Por último, el contexto permite evaluar si la inferencia realizada es coherente a la situación problema que busca explicar.

En síntesis, el RII permite hacer juicios de valor o predicciones basados en muestras; razonar sobre las posibles características de una población basados en una muestra, o sobre cuáles características de las muestras observadas se presenten con mayor probabilidad en la población; y articular los datos y el contexto de los datos como evidencia con las generalizaciones o hipótesis realizadas.

Comprensión, Interpretación y argumentación de la información estadística (CIA)

Interpretar el significado de los elementos estadísticos y proporcionar argumentos basados en datos para argumentar conclusiones, deben ser habilidades que un ciudadano estadísticamente culto desarrolla para consumir responsablemente la información estadística. Es decir, debe estar en la capacidad de leer, interpretar, razonar y evaluar la información estadística sin importar su representación.

La información que se comprenda e interprete puede ser de diferente orden de complejidad, tipo de escritura o de jerga. La interpretación de los datos se da en función de la vida cotidiana y del enriquecimiento del conocimiento de significados de términos estadísticos presentes en los medios de comunicación (v.g. aleatoriedad, azar, porcentaje, promedio).

Según Watson y Moritz (2000; citado por Contreras y Molina, 2019) para que una persona sea estadísticamente culta debe aprender a interpretar la terminología estadística, comprender el lenguaje y conceptos estadísticos y contradecir los reportes cuyas bases no son bien fundamentadas. Se añade lo anterior la postura de Abelson (2012; citado por Contreras y Molina, 2019) el cual propone dos dimensiones para el ejercicio de la argumentación informal; la interpretativa, donde concluye sobre los datos y la retórica narrativa, en donde propone argumentos convincentes con base en el análisis de datos.

Actitud crítica y cuestionadora (AC y C)

Cuestionar y criticar la información estadística permite conocer la fiabilidad de la información fundamentada en la veracidad de los datos, además de encontrar conclusiones más allá de la interpretación de contenido representado y poder reflexionar sobre las posibles interpretaciones alternativas a las conclusiones que sean comunicadas (Contreras y Molina, 2019).

El razonamiento implicado en el cuestionamiento debe considerar la visión socio histórica, la ubicación social, la subjetividad, el contexto político del conocimiento y la comprensión del flujo de la información.

Dificultades y obstáculos asociados al aprendizaje y enseñanza de la IEI

La inferencia estadística da a los estudiantes conocimientos para estudiar fenómenos, establecer predicciones sobre el comportamiento de una población y plantear investigaciones en diferentes contextos. Para su enseñanza, el estudiante requiere tener la habilidad de analizar y sistematizar información, interpretar datos, capacidad de plantear hipótesis, correlacionar datos, etc.

Batanero et al. (1993; citado por Sánchez y Ruiz, 2017) considera dos aspectos a reflexionar al momento de enseñar estadística inferencial: la mayoría de las investigaciones en la didáctica de la enseñanza de la estadística se centran en la estadística descriptiva a nivel primario o universitario. Además, las primeras investigaciones en la enseñanza de la estadística se realizaron por psicólogos, no por pedagogos, por lo tanto, estaban asociadas a

los procesos cognitivos relacionados al aprendizaje, omitiendo la reflexión del acto educativo en el aula.

Añadiendo a lo anterior, existen dificultades en la conceptualización de la estadística tales como: la capacidad de interpretar y comprender los datos, en la representación matemática, la comprensión del concepto azar y el razonamiento proporcional. En cuanto a la comprensión de los términos teóricos de la estadística inferencial los estudiantes tienen dificultades para entender el concepto de muestra, tamaño de muestra o muestreo; para identificar los tipos de análisis estadísticos y su aplicación; y para distinguir entre el estadístico y el parámetro.

Errores que se pueden cometer en el aprendizaje de la estadística inferencial informal

Teniendo en cuenta las propuestas para la enseñanza de la estadística inferencial informal mediante la potenciación del RII de Zieffler et al. (2008), García y Sánchez (2014), Contreras, Cañadas, Gea y Arteaga (2013), se consolidan los siguientes errores asociados al aprendizaje de los estudiantes:

1. Cuando se realizan estimaciones y representan gráficamente el conjunto de datos, así como los elementos de las representaciones (medidas de centralidad, medidas de dispersión, forma), leyendo así predicciones basadas en sus ideas preconcebidas.
2. El uso del lenguaje probabilístico se desarticula de las características del contexto bajo el cual se está trabajando, por lo cual el hilo de razonamiento entre lo observado y analizado en la situación problema se quebranta con facilidad.
3. Realizan juicios incorrectos al no ser conscientes de las diferencias de distribuciones entre dos conjuntos de datos, o realizan su juicio con un razonamiento de evidencias mixtas: para un conjunto consideran la información analizada, mientras que para el otro insisten en dar veracidad a las ideas alojadas fuera del conjunto de datos.
4. Utilizan dos tiempos verbales diferentes, presente y futuro imperfecto, en la argumentación de las hipótesis y/o conclusiones.
5. Argumentan la validez de una hipótesis con base en muestreos no correspondientes, es decir, confunde las diferentes distribuciones realizadas cuando las comparan en paralelo.

6. Construyen el ideal de un intervalo de variación en las pruebas de significación alrededor de la moda. Además, desarticulan el lenguaje probabilístico en los argumentos al concluir la veracidad de la hipótesis asegurando alrededor de esta medida.
7. Confunden las medidas de dispersión con las medidas de centralidad, usando sólo las segundas como argumentación de la variabilidad de la distribución.
8. No reconocen la relación existente entre el centro de la distribución y su variabilidad, según el tamaño de la muestra tomada.
9. Desarticulan la investigación realizada del contexto en el que está realizando la inferencia y se limita a argumentar sobre los elementos disciplinares de la estadística.
10. Las ideas relacionadas a la aleatoriedad no se ven argumentadas en las características de la distribución (v.g. formas, centro, variación) y se restringen a los conocimientos del ambiente escolar no asociado a los elementos estadísticos.
11. En la transición de la estadística descriptiva y la estadística inferencial, no reconocen las diferencias entre: muestreo y población, estadístico y parámetro y descripción e inferencia.
12. En las conclusiones sobre las hipótesis relacionadas con el modelo de la población, no argumentan o argumentan incorrectamente sobre su significación. Es decir, no justifican debidamente esta decisión final.
13. Obvian el análisis de datos e insisten únicamente en sus creencias para argumentar sobre las inferencias, generalizaciones y/o conclusiones realizadas. Es decir, realizan una justificación subjetiva, en lugar de una determinista o racional probabilística.
14. No asocian el azar a las nociones de incertidumbre como elemento del análisis de datos, y le dan características de descripción de la población, obviando así la naturaleza del muestreo aleatorio.
15. Se limitan a realizar la lectura de datos y más allá de los datos (hipótesis, inferencias o generalizaciones). No realizan una lectura entre datos en la construcción de los argumentos, dando juicios poco efectivos.
16. Asocian porcentajes poco creíbles o significativos a las hipótesis, generalizaciones y/o conclusiones.

Valoración del aprendizaje en la estadística inferencial informal

Para valorar el progreso de los estudiantes en este razonamiento, se tiene en cuenta el desarrollo articulado de los tres componentes de la estadística inferencial informal. Primero, sobre las generalizaciones utilizadas para identificar patrones en los datos incluyendo predicciones, estimación de parámetros y conclusiones que van más allá de describir los datos. Segundo, se reconoce la evidencia de los datos que soportan la generalización, predicciones o conclusiones; es decir, los argumentos implícitos o explícitos que son aceptables en el contexto real y justifican las decisiones tomadas. Por último, el uso del lenguaje probabilístico o de incertidumbre, mediante los verbos que soportan las evidencias relacionadas con la incertidumbre (Sánchez y Ruiz, 2017).

Marco metodológico

Contexto y estudiantes participantes

Originalmente, el proyecto se inició en la Institución Educativa Distrital (I.E.D) Colegio Isabel II, ubicada en el sector de Corabastos en la localidad de Kennedy, el cual se caracteriza por ser uno de los comercios abiertos más importantes de Bogotá, tener una predominancia del estrato 2 y en la cual se han dado a conocer en los medios varios casos de acoso. Concibe al estudiante como constructor de su conocimiento bajo una actitud crítica, mediante actividades intencionales, planificadas y sistemáticas, reconociéndolo como ser activo, crítico y consciente de su realidad y la de los que lo rodean (Colegio Isabel II, 2019).

En esta institución participaron ocho estudiantes de grado décimo, bajo la autorización de la directora del área de matemáticas en modalidad virtual y atendiendo a las recomendaciones dadas por el MEN para la continuación de las actividades académicas para mediados del 2021 surgidas por la pandemia del COVID 19. Los estudiantes anteriores eligieron voluntariamente hacer parte del proyecto.

Luego, se les notificó a los padres de familia de los estudiantes mediante el consentimiento informado sobre la intencionalidad del proyecto, solicitando en el mismo la autorización para realizar la toma de evidencias. Los estudiantes entregaron los consentimientos diligenciados y los subieron a la plataforma de Google Classroom que se había designado para el proyecto. Infortunadamente, debido a que el diseño de las actividades tomó más tiempo del esperado, no se pudo continuar el proyecto en modalidad virtual, ya que para ese tiempo los estudiantes estaban asistiendo a la Institución de manera presencial. En consecuencia, la directora del área asignó uno de los cuatro cursos de décimo para aplicar la secuencia de actividades y tareas diseñadas a la luz del análisis de los resultados de la prueba diagnóstica realizada a los ocho estudiantes que estaban participando en el proyecto.

Por otra parte, a pesar de solicitar el aula de informática y las tabletas para que los estudiantes pudieran explorar las actividades propuestas en las guías con el soporte del Excel programado, la Institución no lo permitió. Por lo tanto, se realizaron intervenciones de manera remota vía Google Meet, de tal forma que en el televisor se proyectaba la videollamada recibida desde el computador de la directora del área. En consecuencia, la comunicación con los estudiantes no fue la ideal, tampoco se pudo observar el avance o las

respuestas de los estudiantes en tiempo real, ni atenderlos por grupos como se tenía pensado inicialmente. Además, solo se pudieron ocupar dos sesiones de clase, ya que, por órdenes de la administración del colegio, los estudiantes debían asistir a reuniones de área, izadas de bandera y otros eventos los cuales se llevaron a cabo en estos espacios. En consecuencia, los resultados de este grupo de estudiantes solo se tuvieron en cuenta para mejorar las actividades y tareas propuestas en las guías, mas no para el análisis de los resultados de los estudiantes. En este orden de ideas, se tomó la decisión de realizar la aplicación de la secuencia de actividades y tareas en otra Institución Educativa cuyo contexto y filosofía sobre el perfil de los estudiantes fueran similares. En este sentido, la aplicación de la secuencia tuvo lugar en la I.E.D Nuevo Chile, ubicada en el barrio Olarte de la localidad de Bosa. En esta zona, en la cual predomina el estrato 2, se realizan actividades comerciales y académicas de domingo a domingo en los diferentes locales comerciales, en la casa de la cultura, entre otros. Además, tal Institución busca que los estudiantes potencien su pensamiento en la medida en que asimilan y acomodan los conocimientos adquiridos, siendo capaces de construir sus propios procesos de aprendizaje y de reflexionar sobre la aplicación de estos conocimientos en su diario vivir (I.E.D. Nuevo Chile, 2018). A petición del director del colegio, se le envió un comunicado explicando los objetivos del proyecto y la explícita solicitud de la toma de evidencias. Una vez aprobado el permiso para trabajar con estos nuevos estudiantes, se anunció el comunicado a los padres de familia y a los estudiantes del curso asignado, el cual contaba con 28 estudiantes. De este grupo es que se reportan los resultados de la secuencia de actividades y tareas propuestas en las guías.

Prueba diagnóstica

A continuación, se presenta la tipificación y análisis de las preguntas propuestas para identificar el nivel de los componentes de la IEI (G, D, LP), de la cultura estadística de interés (CIA, AC y C), así como los niveles de comprensión de la información (lectura de datos, lectura entre datos y lectura más allá de los datos).

Tabla 1.

Tipificación de las preguntas propuestas en la prueba diagnóstica.

Pregunta	Componente de la IEI	Componente de la cultura estadística	Nivel de lectura
-----------------	-----------------------------	---	-------------------------

1	D	CIA	Lectura de datos
2	D	CIA	Lectura de datos (i) y Lectura entre datos (ii y iii)
3	D	CIA	Lectura entre datos
4	D	AC y C	Lectura entre datos
5	D	CIA	Lectura entre datos
6	D	CIA	Lectura de datos
7	D	CIA	Lectura de datos
8	D	CIA	Lectura entre datos
9	G, D	AC y C	Lectura entre datos
10	G, D, LP	AC y C, CIA	Lectura más allá de los datos
11	G, D, LP	AC y C, CIA	Lectura más allá de los datos

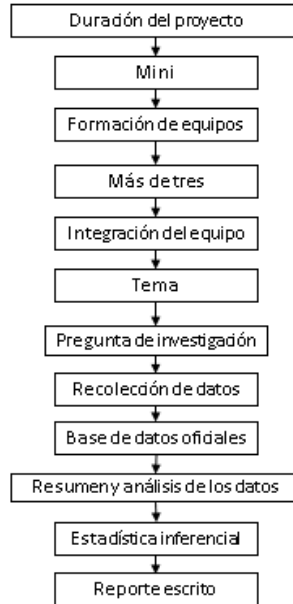
Nota. Actitud crítica y cuestionadora (AC y C); Comprensión, interpretación y argumentación de la información estadística (CIA); Datos como evidencia (D); Generalización (G); Lenguaje probabilístico (LP)

Ruta del proyecto estadístico inferencial informal

En la Figura 6 se resume la ruta a desarrollar en el presente proyecto, adaptada de la descripción realizada por Flores y Pinto (2017), en la cual se especifica aspectos logísticos del proyecto tales como su duración, la formación de los equipos de trabajo y se mencionan los procesos del ciclo investigativo.

Figura 6.

Ruta del proyecto estadístico informal, adaptado de Flores y Pinto (2017)



Propuesta de actividades

A la luz de los resultados obtenidos sobre las habilidades asociadas a los elementos de la cultura estadística de interés a estudiar (comprensión, interpretación y argumentación de la información estadística; actitud crítica y cuestionadora), sobre los niveles de lectura y los componentes de la IEI, se diseñaron tres guías bajo las temáticas de la inferencia de proporciones, inferencia sobre diferencia de proporciones e inferencia sobre medias alrededor del tema de interés seleccionado por los estudiantes: Acoso callejero a las mujeres de Kennedy. La base de datos con la cual se trabajó fue publicada por la Secretaría Distrital de la Mujer (2019) en la página Web <https://datosabiertos.bogota.gov.co/>. Además, se programó un aplicativo en Excel con la base de datos para realizar los diferentes análisis de datos. Este aplicativo estaba programado para que Excel seleccionara datos de la base depurada obtenida de manera aleatoria, en función de la cantidad de habitantes de cada Unidad de Planteamiento Zonal (UPZ).

A continuación, para las preguntas que constituyen las guías se presenta la intencionalidad de cada pregunta y, por último, los componentes del saber disciplinar (cultura estadística y IEI) y del saber matemático.

Guía uno

Inferencia sobre proporciones

Parte 1: Inferencia de proporciones con una muestra de tamaño 10

(plantilla tamaño 10)

1. *Un estudio hecho por la Secretaría de la Mujer reveló alarmantes cifras sobre la consideración de acoso callejero a mujeres en Bogotá. Plantee una afirmación acerca de la proporción (o porcentaje) de mujeres de la localidad de Kennedy que usted piensa que consideran que en la zona ha habido casos de acoso callejero hacia las mujeres. Ejemplos: el 0.4 de las mujeres considerarían que en la zona ha habido casos de acoso callejero hacia las mujeres, más del 40% considerarían que en la zona ha habido casos de acoso callejero hacia las mujeres.*

Intencionalidad: Rastrear indicios sobre la formulación de hipótesis estadísticas relacionadas a la inferencia de proporciones, articulando el lenguaje matemático al contexto de la situación problema.

Para responder las preguntas que siguen, utilice el archivo de Excel que fue preparado para tomar muestra de tamaño 10 en un grupo de 1372 mujeres de la localidad de Kennedy.

Seleccione una muestra de tamaño 10 y responda lo siguiente:

2. *¿Qué piensa acerca del valor de la proporción obtenida con base en esa muestra? ¿Cree realmente que corresponde al que puede tener la población? Explique porque si o porque no.*

Intencionalidad: Utilizar los datos del muestro obtenido para argumentar su respuesta, estableciendo conexiones entre el estimador y el parámetro que se analizan.

3. *Si consideró que la muestra obtenida no representa a la población correctamente, explique qué podría hacer para obtener una muestra significativa.*

Intencionalidad: Proponer ideas con base en la información obtenida para mejorar el resultado obtenido en el muestreo realizado, para discriminar entre el muestreo aleatorio simple y el muestreo estratificado.

4. **Comparar resultados de la muestra obtenida por cada uno con la de otros compañeros.** *Anote los resultados de las proporciones obtenidas con las muestras*

seleccionadas (las suyas y la de sus compañeros) y explique cuál resultado podría estimar mejor el valor de la proporción poblacional que se desconoce.

Intencionalidad: Comparar los datos obtenidos del muestreo individual con el muestro grupal (aquel realizado con los demás integrantes del curso), concluyendo con una actitud crítica y cuestionadora sobre el “mejor” valor posible del parámetro estudiado.

5. *Con base a lo que respondió anteriormente. (i) ¿Qué valores de los estimados son más probables que se presenten en la población? Explique. (ii) ¿En qué elementos de los estudiados (medidas de centralidad, intervalos de dispersión, tablas de proporciones, gráficas de barras) se basó para responder esta pregunta?*

Intencionalidad: (i) Identificar la lectura más allá de los datos, fruto de la relación entre posturas resultantes de las anteriores conclusiones, conectando la argumentación, comprensión e interpretación de los datos obtenidos con el contexto de la situación problema. (ii) Identificar el uso sobre los elementos estudiados en la argumentación de las conclusiones.

Parte 2: Inferencia de proporciones con una muestra de tamaño 30

Seleccione una muestra de tamaño 30 y responda lo siguiente

1. *Con base en la experiencia de lo que vio en la primera parte de esta actividad al tomar varias muestras de tamaño 30, considere la hipótesis acerca de la proporción o porcentaje de mujeres de la localidad de Kennedy que son acosadas en la calle que enunció en la primera actividad. Si cree necesario que la debe cambiar a otro valor, diga a que otro valor. Explique su decisión respecto a conservar o no la hipótesis primeramente enunciada.*

Intencionalidad: Argumentar sobre las decisiones tomadas en torno a la hipótesis planteada anteriormente, tomando como referencia las proporciones obtenidas en el nuevo muestreo.

2. *Organice los resultados de las proporciones obtenidas con las muestras seleccionadas (la suya y la de sus compañeros). Explique cuál resultado cree que está estimando mejor el valor de la verdadera proporción que se desconoce. Para ello utilice información que aportan aspectos tales como las medidas de centralidad, intervalos de dispersión, tablas de proporciones, gráficas de barras, así como sus valores correspondientes.*

Intencionalidad: Argumentar la selección del estimador con base en los elementos obtenidos en el análisis realizado con los datos del grupo.

3. *Explique si cambia de parecer respecto a la validez de la hipótesis enunciada al comienzo.*

Intencionalidad: Decidir cuál de las hipótesis anteriormente propuestas considera mejor para enunciar el estimador, evidenciando así el avance del RII en la construcción de la validez en la argumentación.

Guía dos

Parte 3: Inferencia de proporciones con muestras de tamaño 100 o superior

1. *Plantee una hipótesis sobre la proporción poblacional de las mujeres que son acosadas, con base en una muestra de tamaño 100 e indique el valor de la proporción de la muestra seleccionada.*

Intencionalidad: Proponer el valor del estimador, con base en la lectura de datos dada por el simulador.

2. *Comparta el resultado de la proporción muestral obtenida con la de sus compañeros y observe el gráfico resultante. ¿Cambiaría su hipótesis inicial? ¿Por qué?*

Intencionalidad: Realizar juicios de valor sobre las hipótesis planteadas anteriormente, argumentando sobre el porqué de su elección, a partir de los elementos estudiados.

3. *Con base en las proporciones obtenidas por todo el grupo ¿Cuáles consideran más probables y cuáles menos probables que se presenten en la población? Justifique sus respuestas teniendo en cuenta los elementos que aporta el estudio.*

Intencionalidad: Argumentar las posibles respuestas con base en la información recolectada, usando términos del lenguaje probabilístico.

4. *Concluya una relación entre el tamaño de la muestra y las medidas de dispersión con los intervalos de variación obtenidos en las tres actividades.*

Intencionalidad: Establecer relaciones entre los objetos solicitados, con base en el estudio empírico y analítico de los muestreos realizados.

5. *Seleccione la proporción muestral que mejor pueda representar a la proporción poblacional y valide (argumentar si se rechaza o no) la última hipótesis planteada. ¿Entre qué valores podría estar la proporción poblacional de mujeres que consideran que ha habido casos de acoso callejero hacia las mujeres en la localidad de Kennedy con base en el estudio realizado?*

Intencionalidad: Argumentar juicios de valor sobre la última hipótesis planteada, para reportar una inferencia sobre el parámetro de estudio.

Inferencia de diferencia de proporciones

Parte 1: Inferencia de diferencia de proporciones con muestra de tamaño 30

1. *Considere una muestra de tamaño 30 de mujeres que residen en la UPZ Patio Bonito y las muestras de tamaño relativa en las otras UPZ y halle las proporciones de las encuestadas que están de acuerdo con que en la zona ha habido casos de acoso. Realice una hipótesis de diferencia de proporciones entre (i) Patio Bonito- Américas, (ii) Kennedy Central-Patio Bonito. Ejemplos: la diferencia podría ser mayor que 0, puede no haber diferencia entre las proporciones, la diferencia de proporciones sería de -0,05.*

Intencionalidad: Hacer uso de los resultados dados en la simulación de datos para plantear hipótesis correspondientes a dos diferencias de proporciones, registrando las proporciones indicadas para facilitar la discriminación de la información.

2. *Suponga que para el caso de las diferencias proporcionales entre Kennedy Central y Patio Bonito se ha obtenido una diferencia negativa ¿Qué significa tener una diferencia menor que cero?*

Intencionalidad: Contextualizar la presente norma matemática en la situación problema estudiada, identificando los pre saberes de los estudiantes que justifican esta norma.

3. *Considere una muestra de tamaño 100 para la UPZ Patio Bonito y las muestras relativas de las demás UPZ y halle las diferencias de proporciones solicitadas en el ítem 1. ¿Cambiaría algunas de las hipótesis iniciales? Sí/ No ¿Por qué?*

Si respondió Sí, escriba la nueva hipótesis

Intencionalidad: Validar las hipótesis planteadas con base en la interpretación del análisis de resultados obtenidos para una muestra de tamaño mayor y proponer una nueva hipótesis, en caso de ser necesario.

4. *Con base en los resultados obtenidos decida si acepta o rechaza alguna de las hipótesis planteadas sobre las diferencias de las proporciones. Justifique su respuesta.*

Intencionalidad: Establecer juicios de valor sobre las hipótesis realizadas, reportando para cada caso los elementos estudiados que consideraron en este razonamiento.

Suponga que el intervalo entre diferencias de proporciones entre dos UPZ es de $(-0.1, 0.1)$. Explique qué se puede concluir sobre la diferencia de proporciones en la población entre las dos UPZ para este caso.

Intencionalidad: Deducir y contextualizar la existencia del 0 en el intervalo dado y argumentar sobre el sentido que tiene el reportar este tipo de intervalos.

Guía tres

Inferencia sobre las medias poblacionales con una muestra de tamaño 100

Considere como hipótesis la siguiente afirmación:

Las mujeres que habitan en la UPZ Castilla que dicen que en la zona se han presentado casos de acosos hacia las mujeres, tienen una edad promedio mayor o igual que 42 años.

1. *Tome una muestra de tamaño 100, calcule la media aritmética muestral (o promedio) de las edades de las mujeres encuestadas que dicen que las mujeres han sido víctimas de acoso en la zona UPZ Castilla, y reporte este resultado tanto en esta guía, así como a los encargados de la actividad.*
2. *Reporte las medidas de dispersión y los intervalos de variación que se obtienen en el aplicativo, que contiene los resultados de promedios de otras muestras de tamaño 100 de sus compañeros.*

Intencionalidad: Identificar el estimador, las medidas de dispersión y los intervalos de variación requeridos en los resultados dados por el simulador.

3. *Argumente si rechaza o no la hipótesis planteada y comente en que elementos estudiados (frecuencias absolutas o porcentuales de las gráficas, medidas de dispersión, medidas de centralidad, intervalos de variación, etc.) se basó para dar esta validación.*

Intencionalidad: Hacer uso de los resultados obtenidos para tomar decisiones sobre la hipótesis inicial, para evidenciar así el crecimiento en el RII.

4. *Realice una conclusión para el parámetro estudiado, las edades promedio de las mujeres encuestadas que residen en la UPZ Castilla que están de acuerdo con el hecho de que en la zona se han presentado escenas de acoso hacia las mujeres, con base en los resultados obtenidos en el estudio de la muestra. Utilice la información suministrada por los intervalos de variación y por las gráficas ilustradas para justificar dicha conclusión.*

Intencionalidad: Utilizar los datos como evidencia en el planteamiento de una conclusión, utilizando el lenguaje probabilístico adecuado para inferir así sobre el contexto de la situación problema.

5. *Con base en los resultados obtenidos discuta sobre:*

- (i) *¿De qué edad(es) promedio sería menos probable encontrar una mujer que esté de acuerdo con que en el sector las mujeres son víctimas de acoso, y*
(ii) *¿Qué porcentaje de estas mujeres representan?*

Para cada caso, mencione que tuvo en cuenta para dar su respuesta.

Intencionalidad: Interpretar y argumentar sobre la información obtenida en los gráficos o intervalos de variación para inferir sobre los parámetros solicitados, promoviendo el desarrollo de RII.

6. *Considere lo aprendido en las guías trabajadas y responda ¿Cómo se puede mejorar la calidad del estimador, considerando el tamaño de muestra?*

Intencionalidad: Identificar la relación existente entre el rango o los extremos de intervalos de variación con relación al tamaño de muestra y proponer el muestreo de tal forma que dichos rangos sean reducidos.

7. *Suponga que una mujer va de compras cerca al barrio Portal de las Américas, el cual está ubicado en la UPZ # 46: Castilla. Si las distribuciones de las edades de las*

encuestadas y de las proporciones en la localidad de Kennedy se comportan similares a los de la muestra de tamaño 100 estudiadas. ¿De qué edad sería más probable que la mujer sea acosada? ¿Cuál sería la probabilidad?

Intencionalidad: Hacer uso de la IEI y de las habilidades involucradas en el RII para inferir sobre los parámetros propuestos, realizando una lectura de, entre y más allá de los datos.

Tipificación de las preguntas de la guía propuestas según los componentes del saber disciplinar y del saber matemático

En este apartado se presentan los componentes del saber disciplinar y del saber matemático comprendidos en cada una de las preguntas.

Tabla 2

Componentes del saber pedagógico y disciplinar

Guía 1		
Inferencia sobre proporciones		
Parte 1: Inferencia de proporciones con una muestra de tamaño 10 (plantilla tamaño 10)		
Pregunta	Componentes saber pedagógico	Componentes saber disciplinar
1	G, LP, CIA, AC y C	M1, M2, M3, M5
2	D, AC y C	M2, M3, M5, M6
3	G, D, AC y C	M2, M3
4	G, D, LP, CIA, AC y C	M3, M5
5	G, D, LP, AC y C	M3, M4, M5, M6, M7, M11
Parte 2: Inferencia de proporciones con una muestra de tamaño 30 (plantilla tamaño 30)		
Pregunta	Componentes saber pedagógico	Componentes saber disciplinar
1	G, D, AC y C	M1, M2, M3, M5
2	G, D, LP, AC y C	M5, M6, M7, M8, M9, M11
3	G, D, AC y C, CIA	M3
Guía 2		
Parte 3: Inferencia de proporciones con una muestra de tamaño 100 (plantilla tamaño 100)		

Pregunta	Componentes saber pedagógico	Componentes saber disciplinar
1	G, D, LP, CIA	M1, M2, M3, M5
2	D, AC y C	M3, M7
3	G, D, LP, AC y C	M3, M7, M9, M11
4	G, D, CIA, AC y C	M2, M9, M11
5	G, D, LP, AC y C	M3, M5, M6
Inferencia de diferencia de proporciones		
Pregunta	Componentes saber pedagógico	Componentes saber disciplinar
1	G, D, LP, CIA	M1, M2, M5, M10
2	AC y C	M10
3	G, D, LP, AC y C	M1, M2, M10
4	G, D, CIA, AC y C	M5, M7, M9, M10
5	G, AC y C	M10
Guía 3		
Inferencia sobre las medias poblacionales con una muestra de tamaño 100		
Pregunta	Componentes saber pedagógico	Componentes saber disciplinar
1	D, CIA	M1, M2, M5, M12
2	D, CIA	M8, M11
3	G, D, LP, AC y C	M5, M7, M9, M10, M11, M12
4	G, D, LP, CIA	M6, M7, M11, M12
5	G, D, LP, CIA, AC y C	M4, M5, M6, M7, M8, M10, M11, M12
6	AC y C	M5, M6

7	G, LP, CIA, AC y C	M4, M5, M7, M8, M12
---	--------------------	---------------------


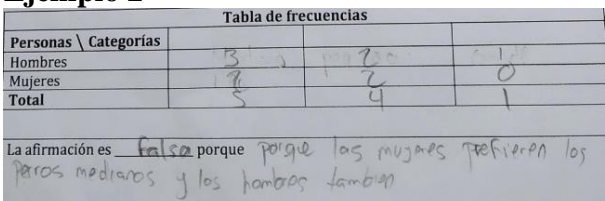
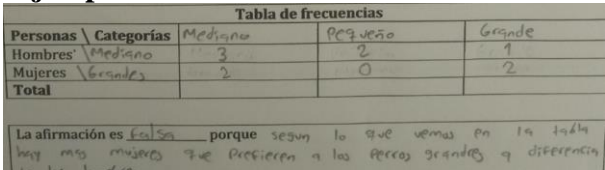
Nota. Actitud crítica y cuestionadora (AC y C); Comprensión, interpretación y argumentación de la información estadística (CIA); Datos como evidencia (D); Generalización (G); Lenguaje probabilístico (LP) M1: Planteamiento de hipótesis estadísticas; M2: Muestreo Aleatorio; M3: Inferencia de proporciones; M4: Proporciones estadísticas; M5: Estadístico; M6: Parámetro; M7: Gráficos estadísticos; M8: Medidas de centralidad; M9: Medidas de dispersión; M10: Inferencia de diferencia de proporciones; M11: Intervalo de variación; M12: Inferencia de medias.

Resultados

Análisis de los resultados de la prueba diagnóstica

En este apartado se presentan los resultados de la aplicación de la prueba diagnóstica (ver el Anexo 1) en la que participaron los ocho estudiantes de grado décimo de la I.E.D Colegio Isabel II. Para cada pregunta se recuerda la intención que tenía y se presenta una breve respuesta esperada a manera de referente para la valoración de los resultados. Además, se agregan algunos ejemplos de respuestas dadas por los estudiantes.

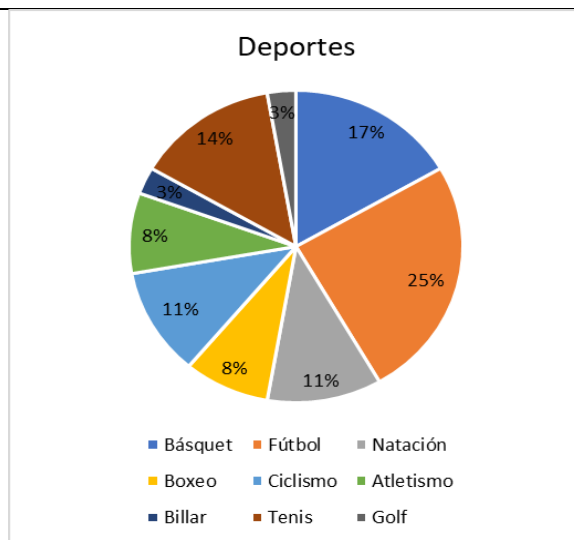
Pregunta 1. Se pide identificar para un grupo de tres tablas de frecuencias dadas, las tres representaciones gráficas que les corresponden de un grupo de cinco.																	
Respuesta esperada	Ejemplos de respuestas dadas por los estudiantes																
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%; text-align: center;">Diagrama de barras</td> <td style="width: 50%; text-align: center;">Tabla de frecuencias</td> </tr> <tr> <td style="text-align: center;">5</td> <td style="text-align: center;">A</td> </tr> <tr> <td style="text-align: center;">4</td> <td style="text-align: center;">B</td> </tr> <tr> <td style="text-align: center;">1</td> <td style="text-align: center;">C</td> </tr> </table>	Diagrama de barras	Tabla de frecuencias	5	A	4	B	1	C	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%; text-align: center;">Diagrama de barras</td> <td style="width: 50%; text-align: center;">Tabla de frecuencias</td> </tr> <tr> <td style="text-align: center;">5</td> <td style="text-align: center;">A</td> </tr> <tr> <td style="text-align: center;">4</td> <td style="text-align: center;">B</td> </tr> <tr> <td style="text-align: center;">1</td> <td style="text-align: center;">C</td> </tr> </table>	Diagrama de barras	Tabla de frecuencias	5	A	4	B	1	C
Diagrama de barras	Tabla de frecuencias																
5	A																
4	B																
1	C																
Diagrama de barras	Tabla de frecuencias																
5	A																
4	B																
1	C																
Resultados. Todos los estudiantes identificaron correctamente los datos representados en los diagramas de barras presentados con sus correspondientes diagramas tabulares.																	
Pregunta 2. Con base en una representación pictográfica se formulan tres preguntas que requieren la habilidad de realizar operaciones aritméticas elementales con los datos dados.																	
Respuestas esperadas	Ejemplos de respuestas dadas por los estudiantes																
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 5px;"> i) ¿Cuántas personas practican Básquet? $10 \cdot 4 = 40$ R//40 </td> </tr> <tr> <td style="padding: 5px;"> ii) ¿Cuál es la diferencia entre las personas que practican Fútbol respecto a las que practican Tenis? R//12 Fútbol: $6 \cdot 4 = 24$ Tenis: $3 \cdot 4 = 12$ Fútbol – Tenis = $24 - 12 = 12$ </td> </tr> <tr> <td style="padding: 5px;"> iii) ¿Cuántos alumnos se encuestaron en total? R//108 estudiantes $27 \cdot 4 = 108$ </td> </tr> </table>	i) ¿Cuántas personas practican Básquet? $10 \cdot 4 = 40$ R//40	ii) ¿Cuál es la diferencia entre las personas que practican Fútbol respecto a las que practican Tenis? R//12 Fútbol: $6 \cdot 4 = 24$ Tenis: $3 \cdot 4 = 12$ Fútbol – Tenis = $24 - 12 = 12$	iii) ¿Cuántos alumnos se encuestaron en total? R//108 estudiantes $27 \cdot 4 = 108$	<div style="border: 1px solid black; padding: 5px; margin-bottom: 10px;"> <p style="text-align: center; margin: 0;">Procedimiento</p> <p>$4 \times 10 = 40$ alumnos</p> <hr/> <p>Fútbol = $6 \times 4 = 24$</p> <p>Tenis = $3 \times 4 = 12$</p> <p>Diferencia de: $24 - 12 = 12$ alumnos</p> <p>$27 \text{ (😊)} \cdot 4 = 108 \leftarrow$ alumnos encuestados</p> </div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 10px;"> <p style="text-align: center; margin: 0;">Procedimiento</p> <p>Lo que se hace es darle valor a cada uno de las caras. $4 \times 10 = 40$</p> <p>Una vez se sabe el número de estudiantes en cada deporte se restan estos números y el resultado será la diferencia. $(24 - 12 = 12)$</p> <p>Aquí se suma el número de estudiantes en cada deporte y el resultado será el total de estudiantes.</p> </div> <div style="border: 1px solid black; padding: 5px;"> <p>ii) ¿Cuál es la diferencia entre las personas que practican Fútbol respecto a las que practican Tenis? R//Los que juegan fut son el doble.</p> </div>													
i) ¿Cuántas personas practican Básquet? $10 \cdot 4 = 40$ R//40																	
ii) ¿Cuál es la diferencia entre las personas que practican Fútbol respecto a las que practican Tenis? R//12 Fútbol: $6 \cdot 4 = 24$ Tenis: $3 \cdot 4 = 12$ Fútbol – Tenis = $24 - 12 = 12$																	
iii) ¿Cuántos alumnos se encuestaron en total? R//108 estudiantes $27 \cdot 4 = 108$																	

<p>Resultados. Todos los estudiantes interpretaron correctamente los pictogramas dados y utilizaron asertivamente la información estadística para responder a las preguntas. Además, varios de los estudiantes (3 de 8) enunciaron con sus propias palabras el procedimiento realizado respecto a la identificación de términos asociados y a los procesos involucrados.</p>																											
<p>Pregunta 3. Se presenta una tabla de registro de datos que contiene los resultados del tamaño de las mascotas que poseen diez personas encuestadas donde se pide registrar totales por tamaño de mascota preferida.</p>																											
Respuesta esperada				Ejemplos de respuestas dadas por los estudiantes																							
Encuestados	Grande	Mediano	Pequeño	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center;">Encuestados</td> <td style="text-align: center;">Grande</td> <td style="text-align: center;">Mediano</td> <td style="text-align: center;">Pequeño</td> <td colspan="2"></td> </tr> <tr> <td style="text-align: center;">Total</td> <td style="text-align: center;">3</td> <td style="text-align: center;">5</td> <td style="text-align: center;">2</td> <td colspan="2"></td> </tr> </table>				Encuestados	Grande	Mediano	Pequeño			Total	3	5	2										
Encuestados	Grande	Mediano	Pequeño																								
Total	3	5	2																								
Total	3	5	2																								
<p>Resultados. Todos los estudiantes reportaron sin alguna dificultad las frecuencias de las variables dadas con base en la información suministrada.</p>																											
<p>Pregunta 4. Se pide con base en la información presentada en la pregunta tres, elaborar una tabla de frecuencias bivariada, totalizar frecuencias para la variable tamaño de los perros y luego valorar la condición de verdad o falsedad de la afirmación presentada (“Las mujeres prefieren los perros pequeños, mientras que los hombres prefieren los perros grandes”)</p>																											
Respuesta esperada				Ejemplos de respuestas dadas por los estudiantes																							
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <th colspan="4" style="text-align: center;">Tabla de frecuencias</th> </tr> <tr> <th style="text-align: left;">Personas \ Categorías</th> <th style="text-align: center;">Mediano</th> <th style="text-align: center;">Grande</th> <th style="text-align: center;">Pequeño</th> </tr> <tr> <td style="text-align: left;">Hombres</td> <td style="text-align: center;">3</td> <td style="text-align: center;">1</td> <td style="text-align: center;">2</td> </tr> <tr> <td style="text-align: left;">Mujeres</td> <td style="text-align: center;">2</td> <td style="text-align: center;">2</td> <td style="text-align: center;">0</td> </tr> <tr> <td style="text-align: left;">Total</td> <td style="text-align: center;">5</td> <td style="text-align: center;">3</td> <td style="text-align: center;">2</td> </tr> </table>				Tabla de frecuencias				Personas \ Categorías	Mediano	Grande	Pequeño	Hombres	3	1	2	Mujeres	2	2	0	Total	5	3	2	<p>Ejemplo 1</p>  <p>La afirmación es <u>Falsa</u> porque la tabla muestra una preferencia por los perros medianos por parte de los hombres y para las mujeres ninguna preferencia hacia los perros pequeños</p>			
Tabla de frecuencias																											
Personas \ Categorías	Mediano	Grande	Pequeño																								
Hombres	3	1	2																								
Mujeres	2	2	0																								
Total	5	3	2																								
<p>La afirmación es falsa porque ninguna mujer prefiere los perros pequeños y son las mujeres quienes prefieren tener perros grandes.</p>				<p>Ejemplo 2</p>  <p>La afirmación es <u>falsa</u> porque porque las mujeres prefieren los perros medianos y los hombres también</p>																							
<p>Ejemplo 3</p>  <p>La afirmación es <u>falsa</u> porque según lo que vemos en la tabla hay más mujeres que prefieren a los perros grandes y diferencia</p>				<p>Ejemplo 4</p> <p>R//La afirmación es falsa porque las mujeres al ser encuestadas prefieren entre un perro mediano y grande mientras que los hombres varían entre pequeño y mediano.</p>																							
<p>Resultados. La mayoría de los estudiantes (5 de 8) identificaron correctamente las variables involucradas para organizar los datos de mayor a menor frecuencia de acuerdo con los totales encontrados para el tamaño de las mascotas, como en el ejemplo 1. No obstante, hubo un estudiante que no organizó las columnas de acuerdo con los totales de la variable tamaño (ejemplo 2); de hecho, no escribe en la tabla dada, los valores de la variable tamaño de las</p>																											

mascotas ya que al parecer, ordena de acuerdo con el género y en un orden descendente, las frecuencias obtenidas. En el ejemplo 3, no se comprende, en principio, porque el estudiante agrega a los valores de la variable género la palabra ‘Mediano’ en hombres y ‘Grande’ en mujeres; quizás está identificando por género la mayor frecuencia de tamaño de mascota. Respecto a los juicios de valor de la afirmación presentada, se evidencia una posición crítica correspondiente concordante con su grado escolar, ya que utilizaron los datos dados para presentar su argumento. En ejemplo 4 se observa que se compara el cada género de acuerdo con la preferencia de mascota.

Pregunta 5. Esta pregunta primeramente pretende que el estudiante realice un proceso de transnumeración de la gráfica presentada en el enunciado, a un diagrama circular o de torta y explicita el procedimiento realizado, y luego, que determine el porcentaje de deportes que no requieren el uso de una pelota.

Respuesta esperada



Básquet: $\frac{6}{36} \cdot 100\% = 17\%$

Fútbol: $\frac{9}{36} \cdot 100\% = 25\%$

Natación: $\frac{4}{36} \cdot 100\% = 11\%$

Boxeo: $\frac{3}{36} \cdot 100\% = 8\%$

Ciclismo: $\frac{4}{36} \cdot 100\% = 11\%$

Atletismo: $\frac{3}{36} \cdot 100\% = 8\%$

Billar: $\frac{1}{36} \cdot 100\% = 3\%$

Tenis: $\frac{5}{36} \cdot 100\% = 14\%$

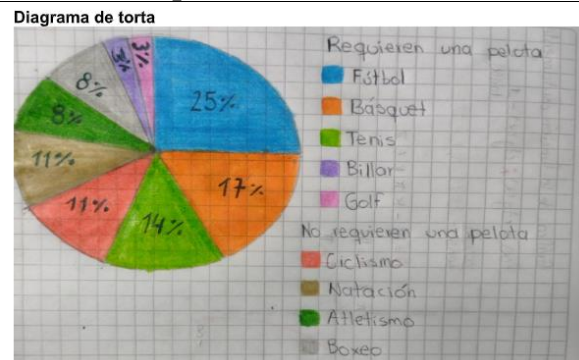
Golf: $\frac{1}{36} \cdot 100\% = 3\%$

Deportes con balón: $\frac{22}{36} \cdot 100\% = 61\%$

Deportes sin balón: $\frac{14}{36} \cdot 100\% = 39\%$

El 39% de los deportes no requieren uso de pelota.

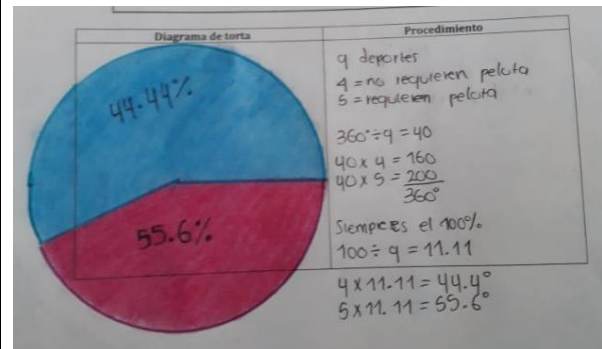
Ejemplos de respuestas dadas por los estudiantes



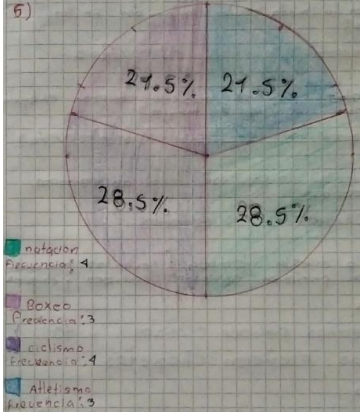
Procedimiento

Deportes	#	Fr	%	360°
Golf	1	1:36 = 0,03	0,03*100 = 3%	0,03*360 = 10,8
Tenis	5	5:36 = 0,14	0,14*100 = 14%	0,14*360 = 50,4
Billar	1	1:36 = 0,03	0,03*100 = 3%	0,03*360 = 10,8
Atletismo	3	3:36 = 0,08	0,08*100 = 8%	0,08*360 = 28,8
Ciclismo	4	4:36 = 0,11	0,11*100 = 11%	0,11*360 = 39,6
Boxeo	3	3:36 = 0,08	0,08*100 = 8%	0,08*360 = 28,8
Natación	4	4:36 = 0,11	0,11*100 = 11%	0,11*360 = 39,6
Fútbol	9	9:36 = 0,25	0,25*100 = 25%	0,25*360 = 90
Básquet	6	6:36 = 0,17	0,17*100 = 17%	0,17*360 = 61,2
	36		100	

Con los resultados del apartado 360° puedo dividir el espacio en el círculo ubicando en sentido de las manecillas del reloj de mayor a menor.

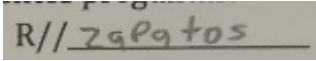
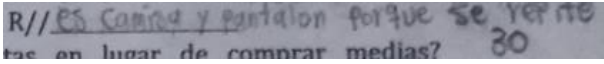
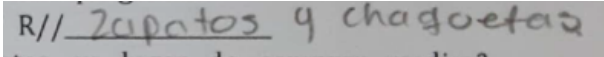


El porcentaje de deportes que no requieren pelota es 33,33%

	
--	--

Resultados. Por un lado, varios estudiantes manifestaron no saber cómo realizar un diagrama circular a partir de los datos dados. Por otro lado, entre los estudiantes que realizaron los diagramas se evidencian múltiples interpretaciones. Hubo quienes construyeron correctamente el diagrama, manifestando un buen nivel de destrezas matemáticas e interpretación lectora; así como también hubo estudiantes que malinterpretaron lo solicitado en la pregunta, construyendo como consecuencia un diagrama circular considerando solo los deportes que no usan balón. También hubo aquellos que obviaron la información del polígono de frecuencias y construyeron el diagrama circular con base en la cantidad de deportes que usan o no alguna pelota.

Pregunta 6. Con esta pregunta se pretende que se indentifique cual es la moda de la prenda que mas compran las mujeres.

Respuesta esperada	Ejemplos de respuestas dadas por los estudiantes
R// Zapatos.	<p>6) ¿Cuál es la moda entre las prendas de las mujeres? R// Camisa y pantalón Procedimiento: Gorra 10, Medias 20, Camisa 30, Pantalón 30, Chaqueta 40, Zapatos 70. El valor que más se repite: 10, 20, 30, 30, 40, 70</p>   

Resultados. Dado que la mayoría de los estudiantes comprendieron a la moda como el valor de frecuencia que se repite, en lugar de la variable con mayor frecuencia, reportaron una respuesta incorrecta. A pesar de ello, se evidencia que hay estudiantes que comprenden correctamente esta medida de tendencia central.

Pregunta 7. Con esta pregunta se busca que el estudiante opere entre datos con base en dos tipos de representación. En efecto se requería que el estudiante calculara el total de datos correspondiente a mujeres, asociara esa cantidad como igual para hombres, determinara la diferencia de porcentajes de tipos de prendas y transformara el porcentaje encontrado en su correspondiente cantidad absoluta.

Respuesta esperada	Ejemplos de respuestas dadas por los estudiantes
--------------------	--

R// el total de mujeres es 200 (suma de frecuencias en el gráfico de mujeres) La diferencia de porcentajes es 20%-15%=5% El 5%, corresponde a $\frac{200 \cdot 5}{100} = 10$ personas.	7) ¿Cuántos hombres prefieren comprar chaquetas en lugar de comprar medias? R// Un 5% prefieren chaquetas en lugar de medias es decir 10 hombres 40 hombres porque: $40/200 = 0.2$ $0.2 \cdot 100 = 20\%$ Chaquetas
--	--

Resultados. Solo un estudiante reportó la respuesta correcta, evidenciando así una apropiada lectura entre los datos dados; un buen nivel de interpretación, comprensión y argumentación de la información estadística; y habilidad para obtener los datos en bruto desde otros sistemas de representación. El resto de los estudiantes que reportaron alguna respuesta para esta pregunta, respondieron en su lugar, la cantidad de hombres que prefieren comprar chaquetas.


Pregunta 8. Se buscaba que los estudiantes realizaran una lectura entre los datos para hallar la cantidad total de personas encuestadas, a partir de la premisa: la cantidad de hombres encuestados fue la misma cantidad de mujeres encuestadas. Además, los estudiantes debían transformar los datos en bruto desde el diagrama circular, para hallar la frecuencia absoluta de las variables observadas.

Respuesta esperada	Ejemplos de respuestas dadas por los estudiantes																					
Mujeres: $30 + 40 + 70 + 20 + 30 + 10 = 200$ Hombres = mujeres = 200 Mujeres + hombres = $200 + 200 = 400$	<p>8) ¿Cuál fue la cantidad total de personas encuestadas? R// 400 personas</p> <p>Procedimiento:</p> <table border="0"> <tr> <td>Mujeres</td> <td>400 personas</td> <td>Hombres</td> </tr> <tr> <td>70 Prefieren zapatos</td> <td></td> <td>50 Prefieren zapatos</td> </tr> <tr> <td>40 Prefieren chaqueta</td> <td></td> <td>40 Prefieren chaqueta 200</td> </tr> <tr> <td>30 Prefieren camisa</td> <td>200</td> <td>30 Prefieren pantalón</td> </tr> <tr> <td>30 Prefieren pantalón</td> <td></td> <td>30 Prefieren camisa</td> </tr> <tr> <td>20 Prefieren medias</td> <td>Mujeres</td> <td>30 Prefieren medias</td> </tr> <tr> <td>+10 Prefieren gorra</td> <td></td> <td>14 Prefieren gorra</td> </tr> </table> <p>200 mujeres ? = hombres</p>	Mujeres	400 personas	Hombres	70 Prefieren zapatos		50 Prefieren zapatos	40 Prefieren chaqueta		40 Prefieren chaqueta 200	30 Prefieren camisa	200	30 Prefieren pantalón	30 Prefieren pantalón		30 Prefieren camisa	20 Prefieren medias	Mujeres	30 Prefieren medias	+10 Prefieren gorra		14 Prefieren gorra
Mujeres	400 personas	Hombres																				
70 Prefieren zapatos		50 Prefieren zapatos																				
40 Prefieren chaqueta		40 Prefieren chaqueta 200																				
30 Prefieren camisa	200	30 Prefieren pantalón																				
30 Prefieren pantalón		30 Prefieren camisa																				
20 Prefieren medias	Mujeres	30 Prefieren medias																				
+10 Prefieren gorra		14 Prefieren gorra																				

Resultados. La mitad de los estudiantes reportaron una respuesta correcta con su debido procedimiento, lo que evidencia habilidad sobre la lectura entre datos, comprensión de la información estadística presentada y razonamiento estadístico. No obstante, uno de los estudiantes a pesar de realizar una asertiva lectura de los datos no tuvo el nivel de habilidad suficiente para hallar la cantidad de hombres encuestados con base en la información reportada hasta el momento. Además, hubo estudiantes que manifestaron tener dificultades inclusive en la lectura de los datos, al reportar como el total de los encuestados, la cantidad correspondiente a la moda de las prendas en las mujeres.

Pregunta 9. Se buscaba que los estudiantes calcularan el total de personas encuestadas que preferirían chaquetas o camisas. Para ello debían obtener las frecuencias absolutas de estas variables en la categoría de los hombres y sumarlas con las frecuencias absolutas correspondientes en la categoría de las mujeres.

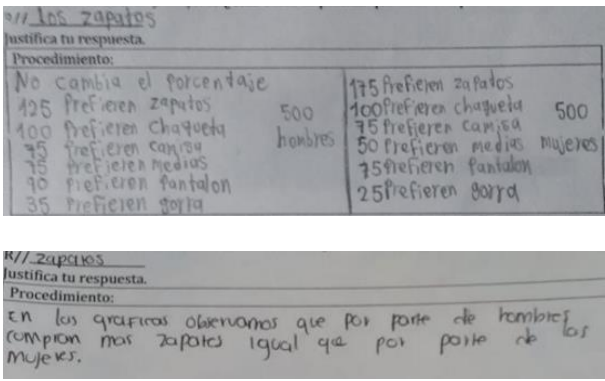
Respuesta esperada	Ejemplos de respuestas dadas por los estudiantes						
Hombres que prefieren camisa: $200 \cdot 15\% = 30$ Mujeres que prefieren camisa: 30 Total: $30 + 30 = 60$ Hombres que prefieren chaqueta: $200 \cdot 20\% = 40$ Mujeres que prefieren chaqueta: 40 Total: $40 + 40 = 80$ Suma pedida: $60 + 80 = 140$	<p>9) En la muestra tomada, ¿Cuántas personas en total prefieren comprar camisas o chaquetas? R// 140 personas</p> <p>Procedimiento:</p> <table border="0"> <tr> <td>Mujeres</td> <td>Hombres</td> </tr> <tr> <td>40 prefieren chaqueta</td> <td>40 prefieren chaqueta</td> </tr> <tr> <td>30 prefieren camisa</td> <td>30 prefieren camisa</td> </tr> </table> <p>70 mujeres + 70 hombres 140 personas = 35% de 400</p>	Mujeres	Hombres	40 prefieren chaqueta	40 prefieren chaqueta	30 prefieren camisa	30 prefieren camisa
Mujeres	Hombres						
40 prefieren chaqueta	40 prefieren chaqueta						
30 prefieren camisa	30 prefieren camisa						

	Camisas Mujeres: $30/200 = 0.15$ $0.15 \cdot 100 = 15\%$ Hombres: $30/200 = 0.15$ $0.15 \cdot 100 = 15\%$	
	Chaquetas Mujeres: $40/200 = 0.2$ $0.2 \cdot 100 = 20\%$ Hombres: $40/200 = 0.2$ $0.2 \cdot 100 = 20\%$	

Resultados. Solo uno de los estudiantes evaluados reportó el total de personas solicitadas. Hubo también quienes reportaron las magnitudes de las variables involucradas sin realizar la suma. Para ambos casos se evidencian habilidades relacionadas a la lectura de datos (extracción de los datos directamente del diagrama de barras) y a la lectura entre datos (cálculo de las frecuencias del diagrama circular).

No obstante, hubo estudiantes que solo resolvieron la pregunta considerando la información de las mujeres o bien, el total de personas que prefieren usar camisa. Es decir, no identificaron correctamente la información estadística que debían considerar.

Pregunta 10. Con el fin de rastrear indicios de una lectura más allá de los datos utilizando los datos de la muestra para realizar inferencias sobre la población, se propuso a los estudiantes estimar el valor de las frecuencias absolutas de cada una de las variables para compararlas y establecer así la moda.

Respuesta esperada				Ejemplos de respuestas dadas por los estudiantes
Prenda	Hombres	Mujeres	Total	
Camisa	$500 \cdot 15\% = 75$	$500 \cdot \frac{30}{200} = 75$	$75 + 75 = 150$	
Chaqueta	$500 \cdot 20\% = 100$	$500 \cdot \frac{40}{200} = 100$	$100 + 100 = 200$	
Zapatos	$500 \cdot 25\% = 125$	$500 \cdot \frac{70}{200} = 175$	$125 + 175 = 300$	
Medias	$500 \cdot 15\% = 75$	$500 \cdot \frac{20}{200} = 50$	$75 + 50 = 125$	
Pantalón	$500 \cdot 18\% = 90$	$500 \cdot \frac{30}{200} = 75$	$90 + 75 = 165$	
Gorra	$500 \cdot 7\% = 75$	$500 \cdot \frac{10}{200} = 25$	$75 + 25 = 100$	
<p><u>Otra opción:</u> observar las gráficas dadas y reportar la variable con mayor frecuencia. R// Zapatos</p>				

Resultados. La mayoría de los estudiantes que respondieron la pregunta manifestaron saber utilizar los datos dados como evidencia para argumentar la hipótesis realizada, utilizando para ello términos representativos del lenguaje probabilístico. Se añade a lo anterior la destreza matemática involucrada en el tránsito entre representaciones que evidenció uno de los estudiantes, para realizar la equivalencia de las frecuencias con este tamaño de muestra.

Pregunta 11. Con el fin de evaluar el nivel de lectura más allá de los datos, se propuso el cambio de las frecuencias relativas en la población para que los estudiantes realizaran el cálculo de las nuevas frecuencias absolutas y, a partir de éstos resultados, indicar la variable que le corresponde la menor frecuencia.

Respuesta esperada	Ejemplos de respuestas dadas por los estudiantes
--------------------	--

R// Gorra.

Suponer una muestra grande con $n = 1000$.

Hombres: $1000 \cdot 70\% = 700$; mujeres:

$1000 \cdot 30\% = 300$.

Prenda	Hombres	Mujeres	Total
Camisa	$700 \cdot 15\% = 105$	$300 \cdot \frac{30}{200} = 45$	$105 + 45 = 150$
Chaqueta	$700 \cdot 20\% = 140$	$300 \cdot \frac{40}{200} = 60$	$140 + 60 = 200$
Zapatos	$700 \cdot 25\% = 175$	$300 \cdot \frac{70}{200} = 105$	$175 + 105 = 280$
Medias	$700 \cdot 15\% = 105$	$300 \cdot \frac{20}{200} = 30$	$105 + 30 = 135$
Pantalón	$700 \cdot 18\% = 126$	$300 \cdot \frac{30}{200} = 45$	$126 + 45 = 171$
Gorra	$700 \cdot 7\% = 49$	$300 \cdot \frac{10}{200} = 15$	$49 + 15 = 64$

Otra opción: observar las gráficas dadas y reportar la variable con menor frecuencia.

R// La Gorra
Justifica tu respuesta.

Procedimiento:

1000 personas	175 P. Zapatos	105 P. Zapatos
70%	140 P. Chaqueta	60 P. Chaqueta
700 Hombres	126 P. Pantalón	45 P. Camisa
30%	105 P. Camisa	45 P. Pantalón
300 Mujeres	105 P. Medias	30 P. Medias
	49 P. Gorra	15 P. Gorra

Procedimiento:

porque en este momento los hombres han comprado un 70% y los mujeres lo mucho menos que las otras prendas

R// La Gorra, porque gracias a ambas graficas nos muestra que tanto mujeres como hombres lo que menos compran son gorras y así nos den que el 70% son hombres o así fueran mujeres ya se sabe con las graficas lo que cada prenda se frecuenta más.

Resultados. Se comprueban indicios de planteamiento de hipótesis estadísticas en las cuales los estudiantes hacen uso de los datos como evidencia, que sugieren una correcta lectura de, entre y más allá de los datos. Además, se observa una actitud crítica y cuestionadora sobre las proporciones dadas en el enunciado para la argumentación de la respuesta.

Análisis de la aplicación de la secuencia didáctica

La aplicación de las actividades y tareas propuestas en la secuencia didáctica tuvo lugar con los 28 estudiantes del grado décimo de la I.E.D. Nuevo Chile. La dinámica general de las sesiones consistió en que uno de los encargados de la actividad dirigía la lectura de las preguntas (algunas veces las leían los estudiantes, otras veces el(la) encargado(a)). Si los estudiantes tenían preguntas sobre la orientación, levantaban la mano, comentaban sus inquietudes y estas eran contestadas. Luego de dar la orientación, los encargados pasaban por los grupos de trabajo atendiendo las diferentes necesidades de aprendizaje, solucionando inquietudes y corroborando que todos estuvieran trabajando en las guías. Al finalizar cada punto, el(la) encargado(a) de dirigir la sesión del día socializaba los resultados observados, dando unas observaciones generales (sugerencias de correcciones a la interpretación de datos o elementos de la estadística trabajados, recalcar los argumentos correctos y/o corrección de

errores) para así continuar con la secuencia. Por último, una vez acabada la sesión se recogían los registros escritos de los estudiantes y se les agradecía por su participación.

La secuencia de actividades propuesta se aplicó en cuatro sesiones de 90 minutos cada una. En la primera sesión los encargados presentaron una breve introducción a los estudiantes sobre la noción básica de la estadística inferencial, estableciendo las diferencias con la estadística descriptiva. Allí se les explicó la diferencia entre muestra y población, la idea de lo que es una hipótesis estadística y la relación entre parámetros y estimadores.

Después, se les presentó el contexto de la situación problema que trata sobre la cantidad de mujeres que consideran que en la localidad de Kennedy han sido víctimas de acoso callejero. Se les preguntó ¿qué entendían por acoso callejero?, ante lo cual algunas de las respuestas estaban asociadas al acoso sexual (eventos en el que los hombres ven con lascivia a las mujeres, les dicen piropos indebidos o intentan tener contacto físico no autorizado). Teniendo estas ideas en consideración, se acordó que el acoso sería entendido como todo acto que pueda maltratar de manera física, emocional o verbal.

Con esto en cuenta, se organizó el grupo de 28 estudiantes en parejas, de tal modo que cada pareja contara con un computador con el aplicativo de Excel y la primera guía orientadora (ver Anexo 2).

Guía uno

En esta sección se comentó a los estudiantes que el objetivo sería inferir o concluir sobre la proporción de mujeres que consideran que han sido víctimas de acoso en la localidad de Kennedy. Para lo cual, se le pidió a uno de los estudiantes que leyera el primer ítem en el que se presentaba la situación problema y la primera indicación:

Parte 1: Inferencia de proporciones con una muestra de tamaño 10 (plantilla tamaño 10)

- 1. Un estudio hecho por la Secretaría de la Mujer reveló alarmantes cifras sobre la consideración de acoso callejero a mujeres en Bogotá. Plantee una afirmación acerca de la proporción (o porcentaje) de mujeres de la localidad de Kennedy que usted piensa que consideran que en la zona ha habido casos de acoso callejero hacia las mujeres. Ejemplos: el 0.4 de las mujeres considerarían que en la zona ha habido casos*

de acoso callejero hacia las mujeres, más del 40% considerarían que en la zona ha habido casos de acoso callejero hacia las mujeres.

Para esta parte fue necesario enunciarles a los estudiantes la manera en cómo se obtienen las proporciones y lo que representan, ya que habían comentado que desconocían este objeto. En respuesta, con la ayuda del aplicativo, se les mostró la manera cómo Excel hacía este cálculo (Figura 7).

Figura 7.

Cálculo de proporciones muestrales en Excel para un tamaño de muestra n=10. Resultados obtenidos con los estudiantes del proyecto, febrero 2022

Individuo	Selección	UPZ	Valor
1	223	Américas	1
2	55	Américas	1
3	276	Castilla	0
4	458	Castilla	1
5	611	Corabastos	1
6	733	Kennedy Central	1
7	704	Kennedy Central	0
8	1107	Patio Bonito	1
9	1257	Patio Bonito	1
10	1067	Patio Bonito	1
Proporción	0.80		
Total	8	Estimado total	1098
Porcentaje	80%		

Se les explicó a los estudiantes que el programa contaba la cantidad de personas que consideran que las mujeres han sido acosadas en la localidad de Kennedy y las dividía entre el total de la muestra. Es decir, que esta proporción representaba la fracción $\frac{\# \text{ personas que cumplen la característica estudiada (variable)}}{\text{total de la muestra}}$. Además, se les explicó que las proporciones se representan con un número decimal del 0 al 1; dónde “0” se podía interpretar como ninguna persona cumple la característica estudiada, y “1” como todas las personas cumplen la característica estudiada. También se podía simbolizar como porcentaje, de 0% al 100%.

Como los estudiantes manifestaron con voz audible que habían entendido la explicación, propusieron la primera hipótesis del estudio. A continuación, se presentan algunas estas hipótesis (Figura 8):

Figura 8.

Algunas hipótesis propuestas por los estudiantes para la muestra de tamaño $n=10$.
Resultados obtenidos con los estudiantes del proyecto, febrero 2022

1. yo creeria que el acoso callejero esta en un 60% en la localidad de Kennedy, tanto en acoso sexual verbal y acoso se apropiacion de ellas y las intentan robar. (a)
El acoso en las mujeres es de un 0,5 (b)
y esta 95% mas o menos de acoso de mujeres en la calle. (c)
1. Hipotesis: Creemos que el 50% de las mujeres en las localidades han sufrido de acoso callejero, Porque creemos que hay la misma igualdad en la lista de gente no acosada y si acosada. (d)

Las hipótesis (a), (c) y (d) de la **¡Error! No se encuentra el origen de la referencia.** dan muestra de los primeros indicios del uso de lenguaje probabilístico mediante expresiones tales como “creería”, “más o menos” y “creemos”. La mayoría de las hipótesis planteadas recalcan la creencia de los estudiantes, ignorando la información dada en Excel o no argumentando su planteamiento (v.g. Figuras 8.b y 8.c). No obstante, sólo una pareja intentó utilizar justificar su propuesta (Figura 8.d).

Con base en los aportes dados por los estudiantes, se les explicó que las hipótesis se deberían argumentar con base en los datos recolectados, porque la idea era inferir sobre el comportamiento de la población, a partir de la muestra obtenida en el aplicativo. Con esto dicho, se dio lugar al segundo punto de la guía:

2. ¿Qué piensa acerca del valor de la proporción obtenida con base en esa muestra?
¿Cree realmente que corresponde al que puede tener la población? Explique porque sí o porque no.

Como los estudiantes manifestaron que no entendían la pregunta ni tampoco lo que es una UPZ, se les explicó que la localidad de Kennedy está dividida en cinco sectores o unidades

por zona (UPZ). En este sentido, se les enunció que este tipo de muestreo es pertinente o es apropiado cuando se puede hablar de todo el conjunto. Como respuesta, algunos estudiantes manifestaron que el resultado del muestreo inicial (ver Figura 9) no estaba bien realizado porque no se podía concluir nada sobre la UPZ Américas (v.g. Figura 10.a). Sin embargo, hubo dos grupos que afirmaron que el muestreo se había hecho de manera correcta (v.g. Figura 10.b).

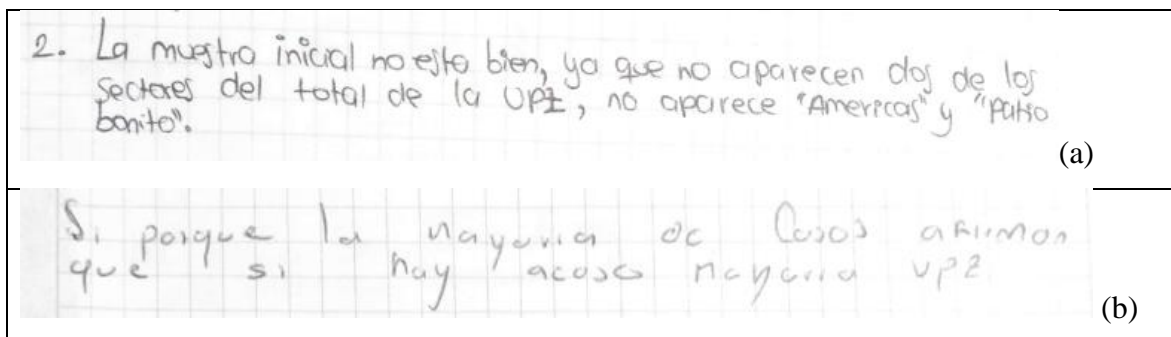
Figura 9.

Ejemplo del muestreo aleatorio simple por UPZ para un tamaño de muestra n=10. Resultados obtenidos con los estudiantes del proyecto, febrero 2022

Muestreo Inicial: n= 10		
Individuo	UPZ	Acoso a mujeres (1: sí; 0: no)
544	Corabastos	1
817	Kennedy Central	0
591	Corabastos	1
842	Kennedy Central	1
741	Kennedy Central	0
450	Castilla	1
763	Kennedy Central	0
358	Castilla	1
564	Corabastos	0
315	Castilla	1
Proporción		0.6

Figura 10.

Argumentos sobre el muestreo inicial para un tamaño de muestra n=10. Resultados obtenidos con los estudiantes del proyecto, febrero 2022



Considerando los aportes dados por los estudiantes, se acordó que el muestreo no estaba bien realizado. Además, uno de los estudiantes explicó con voz audible para aquellos que no habían entendido su error en la respuesta, mencionando que “el muestreo debería tener a gente de Américas y Patio Bonito porque hacen parte de Kennedy”.

Vale la pena resaltar que el formalismo matemático de este tipo de muestreo y el estudio de otras técnicas de muestreo, no eran centro fundamental de los objetos de estudio de la presente secuencia. Por lo tanto, no se entró en detalle sobre la obtención proporcional del tamaño de muestra para cada uno de los estratos.

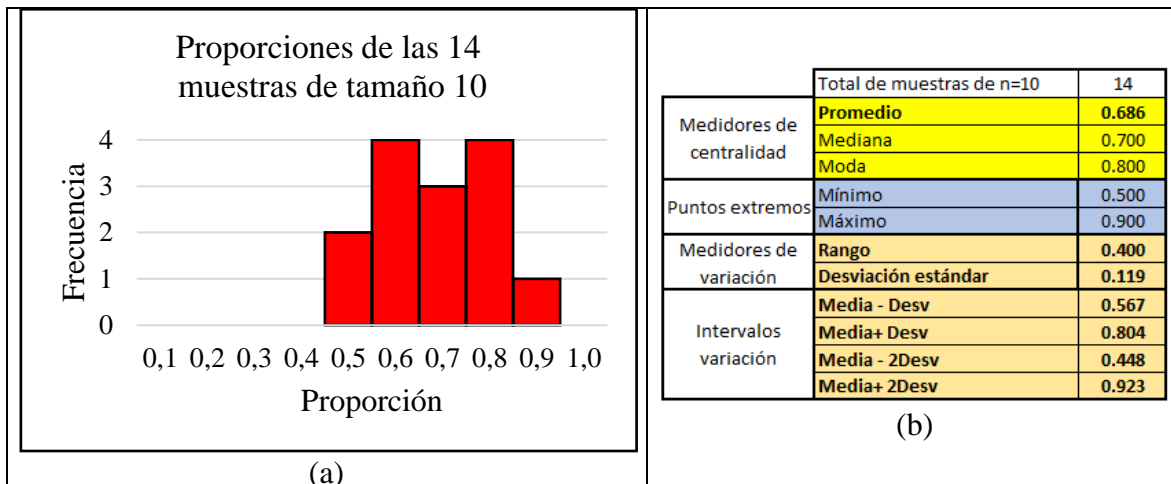
En este orden de ideas, considerando que el tiempo restante no era óptimo para entrar en el detalle de la diferencia entre los muestreos, se decidió omitir el tercer punto y pasar directamente al cuarto punto de la guía. Lo anterior se decidió, porque a la investigación realizada se le dio prioridad al desarrollo de argumentos que dieran cuenta de la formulación y validación de las hipótesis, por encima de la identificación de los tipos de muestreo.

4. Comparar resultados de la muestra obtenida por cada uno con la de otros compañeros. *Anote los resultados de las proporciones obtenidas con las muestras seleccionadas (las suyas y la de sus compañeros) y explique cuál resultado podría estimar mejor el valor de la proporción poblacional que se desconoce.*

Los estudiantes mencionaron con voz audible la proporción dada por el aplicativo; mientras que los encargados de la actividad iban registrando en el Excel estos datos. Se les pidió a los estudiantes que registraran en su archivo de Excel los datos que se estaban ingresando, para luego comparar estos valores y la media de las proporciones con la proporción obtenida en su muestreo. En la Figura 11 se ilustran el gráfico estadístico obtenido, las medidas de centralidad, de dispersión y los intervalos de variación de la distribución de los datos recolectados.

Figura 11.

Gráfico estadístico y tabla resumen del análisis de datos consolidada con las muestras dadas por los estudiantes. Resultados obtenidos con los estudiantes del proyecto, febrero 2022



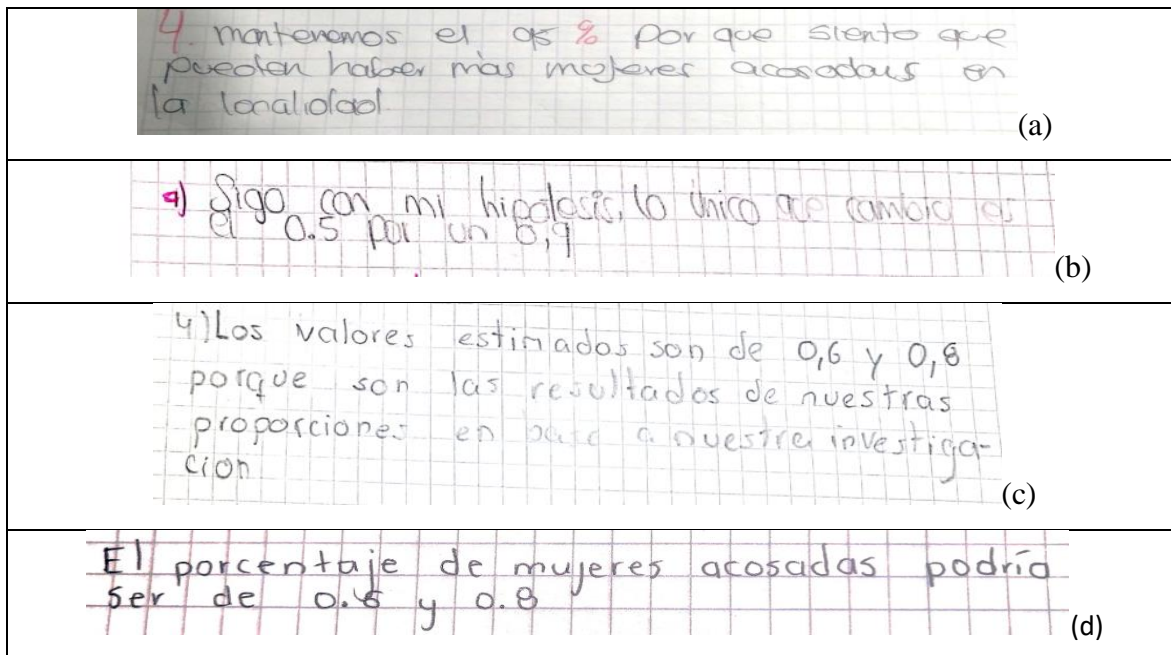
Para este momento, se les recomendó a los estudiantes que tuvieran en cuenta los resultados obtenidos en la gráfica y en la tabla resumen en su decisión, puesto que estas evidencias ayudarían a que su argumento fuera más convincente (Makar & Rubin, 2009). Los conceptos de las medidas de centralidad y el rango ya eran conocidos por los estudiantes, debido a que los habían trabajado en años anteriores. Sin embargo, pidieron que se les explicara los conceptos de desviación estándar e intervalos de variación. De lo anterior, se les comentó que la desviación estándar “nos ayuda a ver qué tanto varía la distribución de los datos alrededor del promedio; mientras más grande sea este dato, más dispersa estará la distribución”.

Para explicar la idea de los intervalos de variación, se localizó en el gráfico estadístico proyectado en el tablero la media aritmética y se encerró en un rectángulo la zona correspondiente a la media aritmética \pm dos desviaciones estándar, indicando que estos valores podrían ser las proporciones correspondientes a la población de mujeres que consideran que en la localidad de Kennedy han sido víctimas de acoso.

Teniendo en cuenta lo anterior, los estudiantes dieron el siguiente tipo de respuestas (Figura 12):

Figura 12.

Respuestas dadas por algunos estudiantes del proyecto a la pregunta 4 sobre la inferencia de proporciones para el conjunto de muestreos consolidados, febrero 2022



Los estudiantes que cambiaron su hipótesis (Figura 12.b) no establecieron argumentos que respaldaran su decisión, y algunos grupos continuaban realizando justificaciones subjetivas (v.g. Figura 12.a). Fueron muy pocos quienes dieron indicios de utilizar la información consolidada como evidencias (v.g. Figuras 12.c y 12.d); sin embargo, uno de los grupos (v.g. Figura 12.c) atribuyeron estos valores a la moda mediante una descripción de los valores, desarticulando el lenguaje probabilístico. Atendiendo a las respuestas que estaba dando el grupo, se vio la necesidad de recalcar que se están realizando inferencias, es decir, que por medio de una muestra se busca concluir sobre el comportamiento de una población. Para ayudar a los estudiantes en el proceso del RII, en particular en la argumentación y validación, se les aclaró que los valores a los cuales les atribuyeron la moda pueden estar presentes en la población, pero también los demás evidenciados en el gráfico estadístico y en la tabla resumen. Claro está, que esos valores podrían encontrarse en menor medida.

Aun evidenciando vacíos conceptuales y procedimentales sobre el planteamiento de las hipótesis estadísticas y la interpretación de los gráficos estadísticos, se tomó la decisión de realizar el muestreo de tamaño 30 con los estudiantes. En ese momento se les sugirió que tuvieran en cuenta la hipótesis planteada en el muestreo de tamaño 10 y la proporción que esta vez les mostraba el aplicativo para el planteamiento de su nueva hipótesis, si decidían cambiarla.

Para ello, se les indicó que fueran a la pestaña llamada “Proporciones n=30” en el aplicativo de Excel.

Parte 2: Inferencia de proporciones con una muestra de tamaño 30

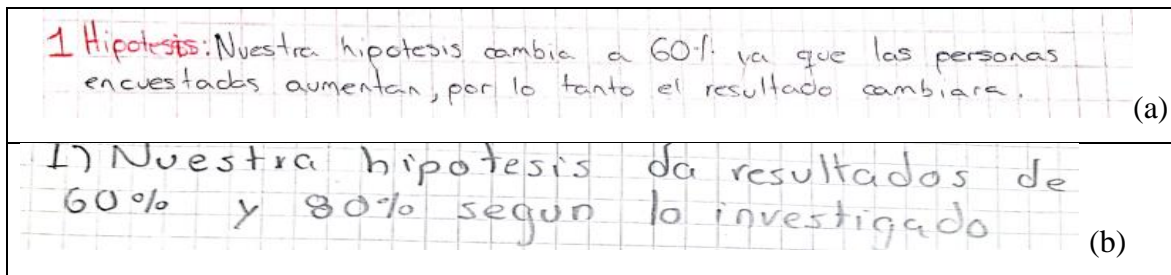
Seleccione una muestra de tamaño 30 y responda lo siguiente

- 1. Con base en la experiencia de lo que vio en la primera parte de esta actividad al tomar varias muestras de tamaño 30, considere la hipótesis acerca de la proporción o porcentaje de mujeres de la localidad de Kennedy que son acosadas en la calle que enunció en la primera actividad. Si cree necesario que la debe cambiar a otro valor, diga a que otro valor. Explique su decisión respecto a conservar o no la hipótesis primeramente enunciada.*

A pesar de las recomendaciones dadas, solo cuatro grupos dieron indicios de haber razonado con base en lo observado en el nuevo muestreo (v.g. Figuras 13.a y 13.b). Sin embargo, para el caso ilustrado en la Figura 13.b, se evidencia falta conceptual en el planteamiento de la hipótesis, ya que la idea se relaciona más con una conclusión. En el caso de los demás estudiantes no se identificó indicio alguno de la argumentación sobre la decisión tomada. Socializando el literal, el encargado de la sesión les recalcó la importancia de hacer uso de la información obtenida y registrarla, puesto que de esta manera se podía dar cuenta del cómo están razonando para plantear las hipótesis y de esta manera, ayudarlos en este proceso. Siguiendo, aclaró que la hipótesis de la Figura 13.b se podía plantear como “la proporción de mujeres que consideran que hay víctimas de acoso en la localidad de Kennedy, podría estar entre 60% y 80%”, haciendo énfasis en que también podían reportar un intervalo de proporciones. Les aclaró también que la hipótesis era una proposición, enunciado u oración realizada a partir de datos, que debería servir para iniciar una investigación o dar una argumentación; adaptando así el formalismo tomado de Mendenhall et al. (2006).

Figura 13.

Ejemplos de hipótesis para una muestra de tamaño n=30. Resultados de los estudiantes del proyecto, febrero 2022



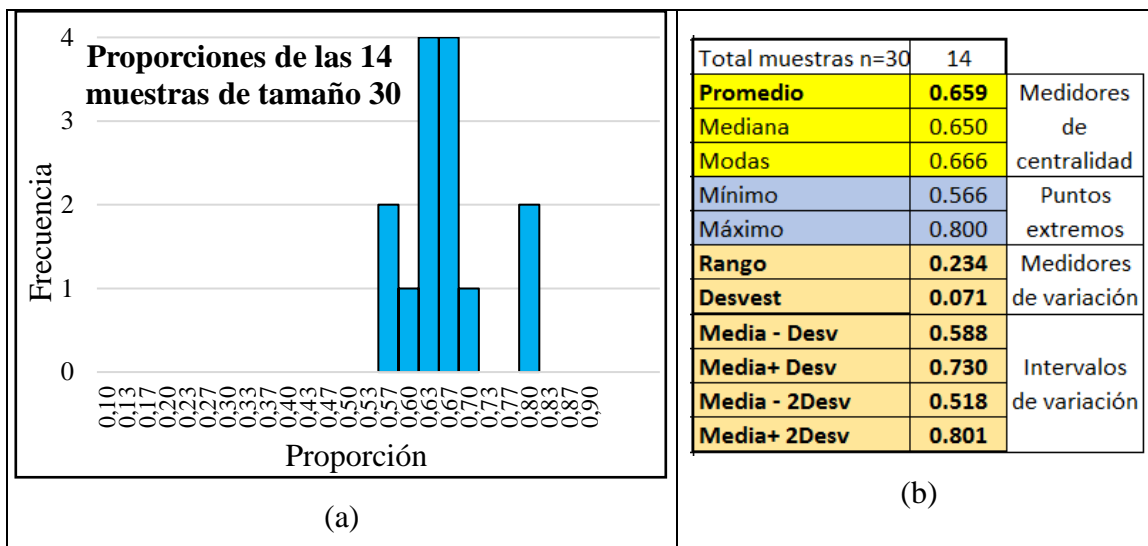
Tras hacer esta aclaración, se dio lugar a realizar el segundo punto de esta parte de la guía, recordando a los estudiantes que el estimador era el dato de interés que están calculando en el muestreo, identificándolo verbalmente por alguno de ellos como la proporción.

2. *Organice los resultados de las proporciones obtenidas con las muestras seleccionadas (la suya y la de sus compañeros). Explique cuál resultado cree que está estimando mejor el valor de la verdadera proporción que se desconoce. Para ello utilice información que aportan aspectos tales como las medidas de centralidad, intervalos de dispersión, tablas de proporciones, gráficas de barras, así como sus valores correspondientes.*

Para cambiar la dinámica, el encargado de la sesión pidió a un integrante por pareja que registrara la proporción obtenida en el aplicativo en el documento de Excel que estaba proyectado. De esta forma, los demás integrantes del grupo registraban en sus aplicativos estos valores, obteniendo así el gráfico estadístico y la tabla resumen correspondientes al muestreo (Figura 14).

Figura 14.

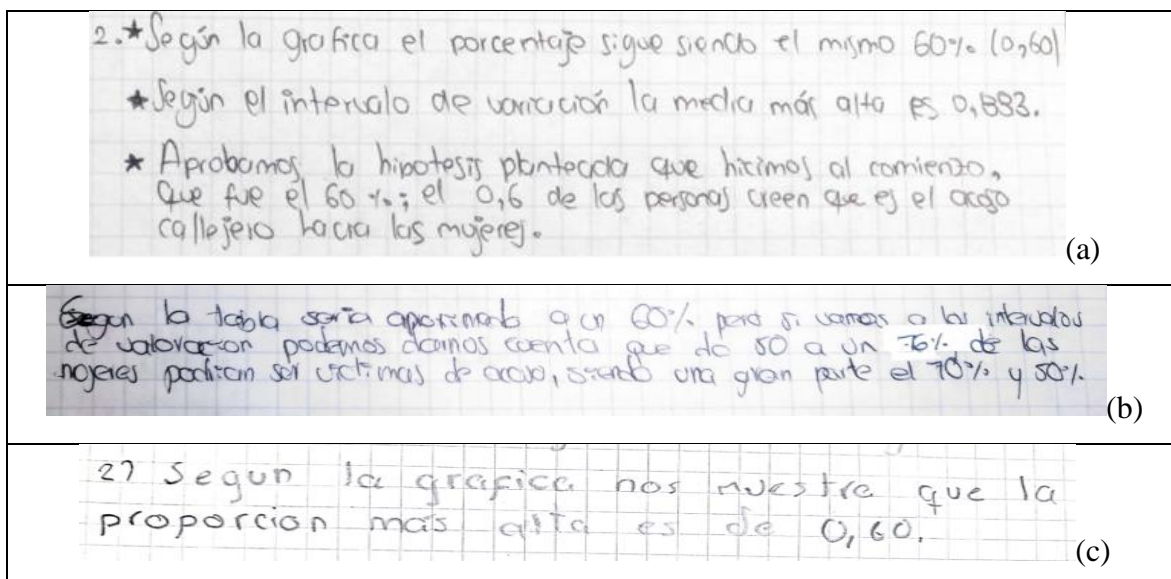
Gráfico estadístico y tabla resumen de las proporciones muestrales consolidadas. Resultados obtenidos con los estudiantes del proyecto, febrero 2022



Considerando la información conjunta, los estudiantes plantearon ideas como las ilustradas en la Figura 15:

Figura 15.

Argumentos sobre la proporción muestral para un muestreo de tamaño $n=30$. Resultados de los estudiantes del proyecto, febrero 2022



Por un lado, se puede evidenciar el avance de los procesos de razonamiento en cuanto la mayoría de los estudiantes reportaron haber usado alguno de los elementos estadísticos estudiados (v.g. intervalo de variación, gráfico estadístico), así como también un avance significativo en la validación y cuestionamiento de las hipótesis (v.g. Figura 15.a). Por otro

lado, aún había estudiantes que no interpretaban correctamente el intervalo de variación, reportando como únicos valores existentes sus extremos. Este error se tomará a colación más adelante, cuando se trate sobre la inferencia de diferencia de proporciones.

Como último literal trabajado en la sesión, se les pidió a los estudiantes que realizaran una decisión final sobre la validación de las hipótesis planteadas en la muestra de tamaño $n=10$.

3. *Explique si cambia de parecer respecto a la validez de la hipótesis enunciada al comienzo.*

Solo un grupo no cambió la hipótesis inicial, ya que desde el primer muestreo el estimador obtenido se acercó bastante a los valores más probables que describirían a la proporción de mujeres en Kennedy que consideran han sido víctimas de acoso. Por último, se consideró para la siguiente sesión darle mayor importancia al contexto de la situación problema, ya que en muy pocos registros se evidencia que los estudiantes lo tengan en cuenta.

Guía dos

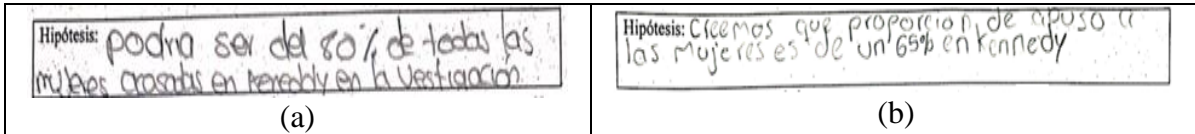
En la segunda sesión se abordó la tercera parte de la inferencia informal acerca de una proporción poblacional, en donde ahora se consideró una muestra de tamaño 100, como se muestra en el siguiente enunciado.

1. *Plantee una hipótesis sobre la proporción poblacional de las mujeres que son acosadas, con base en una muestra de tamaño 100 e indique el valor de la proporción de la muestra seleccionada.*

Los estudiantes plantearon la correspondiente hipótesis, pero solo uno de ellos intentó revisar las evidencias sobre la hipótesis a plantear en sus registros escritos que se obtenían de los muestreos aleatorios que producía el aplicativo (Figura 16.a). Sin embargo, en su discurso verbal el 79% de los grupos reportó que esta población podría estar entre el 60% y 65% (v.g. Figura 16.b) y el 21% restante, reportaron como posibles porcentajes 80%, 85% o 90% sin tener en cuenta los datos que reportaba el aplicativo.

Figura 16.

Hipótesis sobre la proporción muestral para un muestreo de tamaño $n=100$. Resultados de los estudiantes del proyecto, febrero 2022



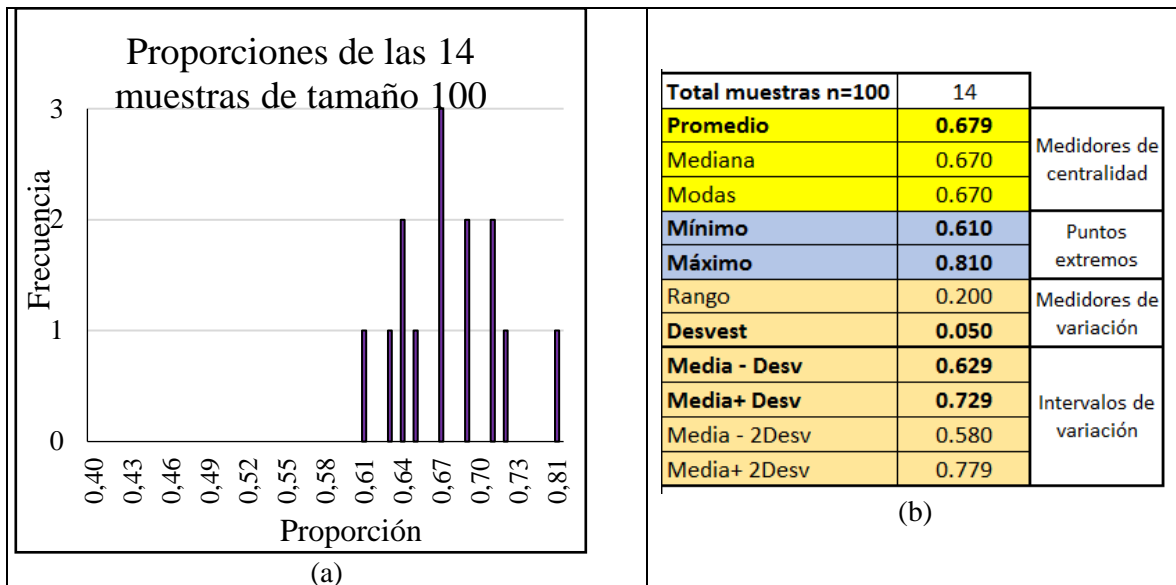
Luego, continuaron con el siguiente punto de la guía:

2. Comparta el resultado de la proporción muestral obtenida con la de sus compañeros y observe el gráfico resultante. ¿Cambiaría su hipótesis inicial? ¿Por qué?

Un miembro de cada grupo digitaba en el Excel proyectado la proporción obtenida; mientras que los demás, ingresaban estos valores en sus aplicativos. A continuación, se muestra el consolidado (Figura 17):

Figura 17.

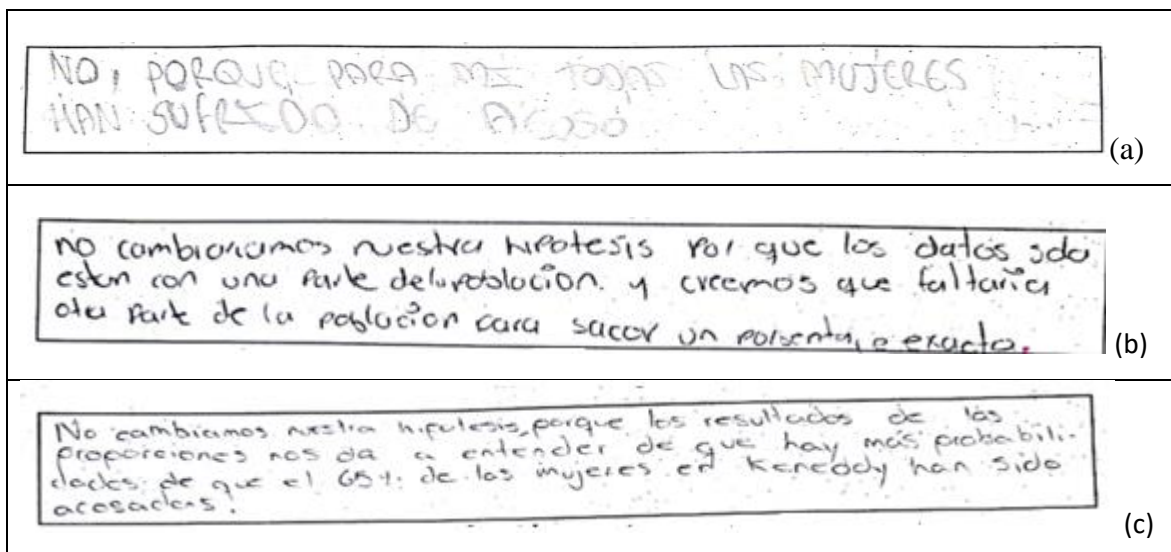
Representación gráfica de la distribución y tabla resumen consolidadas para un muestreo de tamaño $n=100$. Resultados obtenidos con los estudiantes del proyecto, febrero 2022



A pesar de que tuvieron en cuenta el contexto de la situación problema para reportar sus respuestas, aún había estudiantes que realizaban justificaciones subjetivas (v.g. Figura 18a), otros creían que se estaba haciendo un estudio descriptivo (v.g. Figura 18.b). Aun así, el 64% de los argumentaron la validez, referenciando el estudio realizado (v.g. Figura 18.c).

Figura 18.

Validaciones sobre las hipótesis propuestas para una proporción muestral para un muestreo de tamaño $n=100$. Resultados de los estudiantes del proyecto, febrero 2022



Se hizo un llamado a los estudiantes que reportaron justificaciones subjetivas, comentándoles que en el estudio estadístico es necesario argumentar con base en los datos para dar credibilidad a sus argumentos. Para el caso de la respuesta ilustrada en la Figura 18.b, se le preguntó al grupo a qué se referían con “porcentaje exacto”, ante lo cual el grupo argumentó que “el porcentaje exacto es el de la población”; por lo tanto, se les aclaró el propósito de realizar el estudio inferencial: lograr concluir sobre el comportamiento de la población, a partir de una muestra. En cuanto a las demás validaciones (ver Figura 18.c), se les felicitó por haber logrado un avance en la construcción de sus argumentos, utilizando el análisis de datos como evidencia y comunicándolos con terminología del lenguaje probabilístico.

Luego de socializar las validaciones, el grupo continuó con el siguiente ítem de la guía:

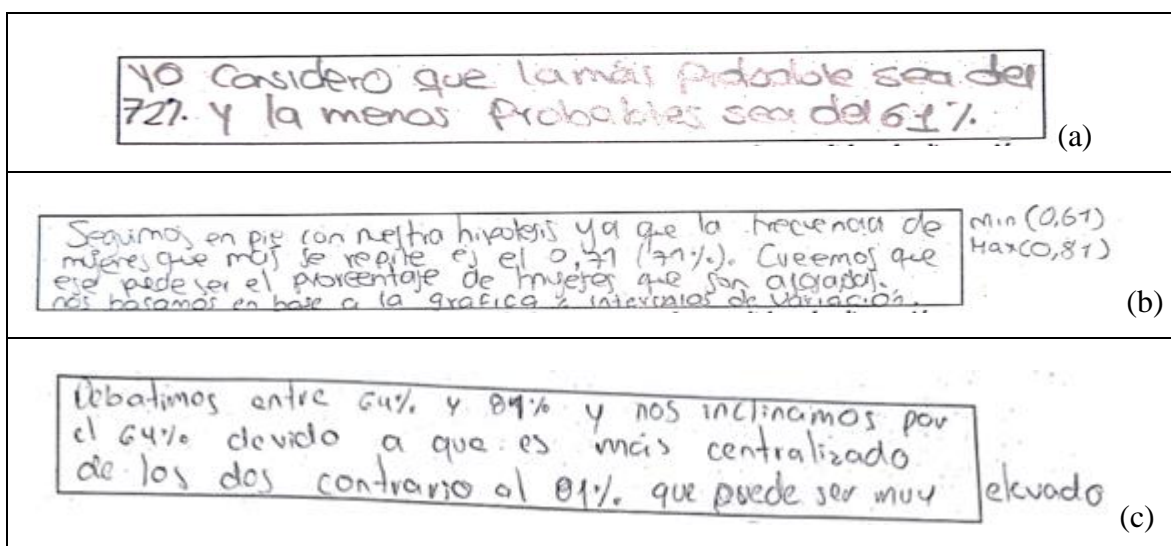
3. Con base en las proporciones obtenidas por todo el grupo ¿Cuáles consideran más probables y cuáles menos probables que se presenten en la población? Justifique su respuesta teniendo en cuenta los elementos que aporta el estudio.

Por un lado, el 50% de los 28 estudiantes localizó las estimaciones pedidas, pero no registraron los elementos tenidos en cuenta para su respuesta (v.g. Figura 19.a). Por otro lado, la otra mitad del curso argumentaron sus conclusiones con base en los intervalos de variación

y/o la gráfica obtenida. Además, se evidencia el desarrollo RII, mediante argumentos basados en los intervalos de variación (Figura 19.b), así como la articulación de conceptos que los estudiantes aprendieron en la estadística descriptiva tales como el centro y variación de la distribución (Figura 19.c).

Figura 19.

Argumentos sobre los posibles valores máximos y mínimos del parámetro. Resultados de los estudiantes del proyecto, febrero 2022



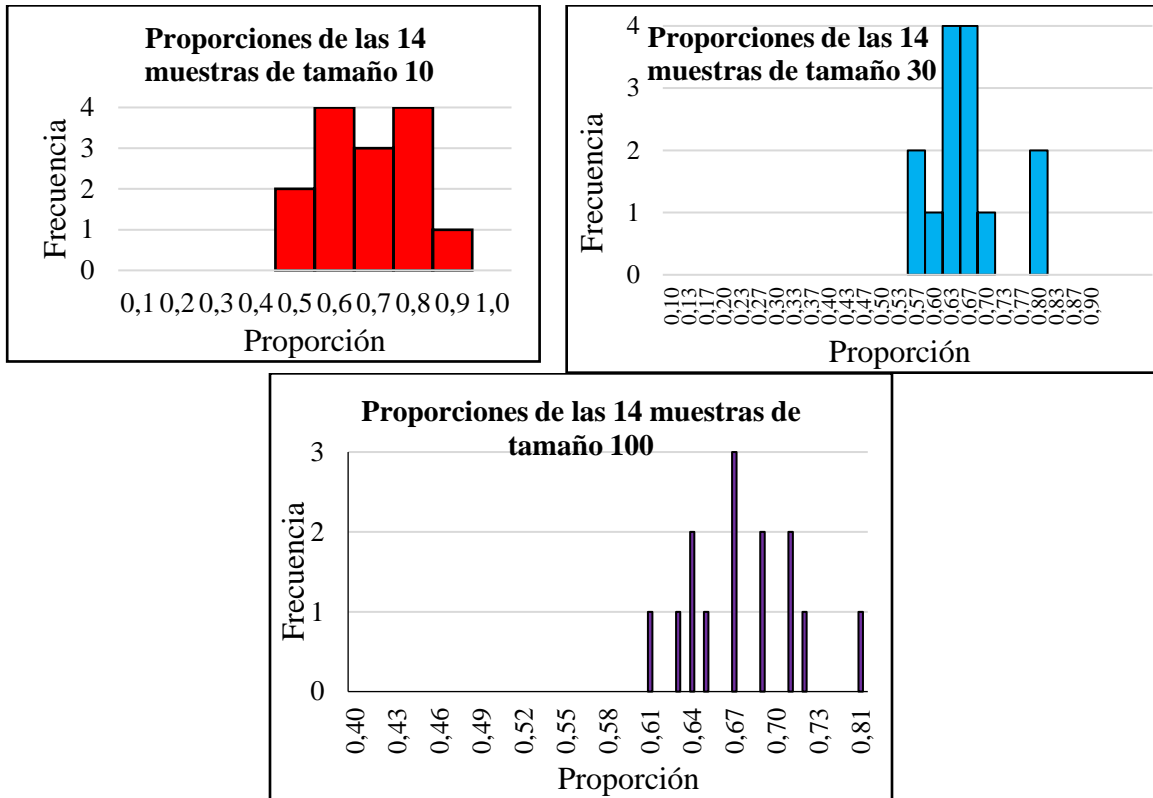
Se recaló la importancia de utilizar los datos como evidencia porque son indispensables al momento de establecer argumentos sólidos sobre una situación, ya que de lo contrario las ideas se quedarían en la subjetividad. En el caso de los demás estudiantes, hubo un avance en la capacidad argumentativa y articulación entre el contexto de la situación problema y los conceptos y procesos estudiados, evidenciando un orden correcto en las ideas del centro de la distribución, según lo enuncian en diversos textos de estadística (v.g. Levine et al. (2006), Anderson et al. (2010) y Mendenhall (2010), entre otros).

Después de socializar el literal, se les mostró a los estudiantes en el Excel proyectado las representaciones gráficas de las tres muestras estudiadas (Figura 20). Se les pidió a los estudiantes que leyeran muy bien las gráficas para poder proponer la relación buscada en la guía.

4. Concluya una relación entre el tamaño de la muestra y las medidas de dispersión e intervalos de variación obtenidos en las tres actividades.

Figura 20.

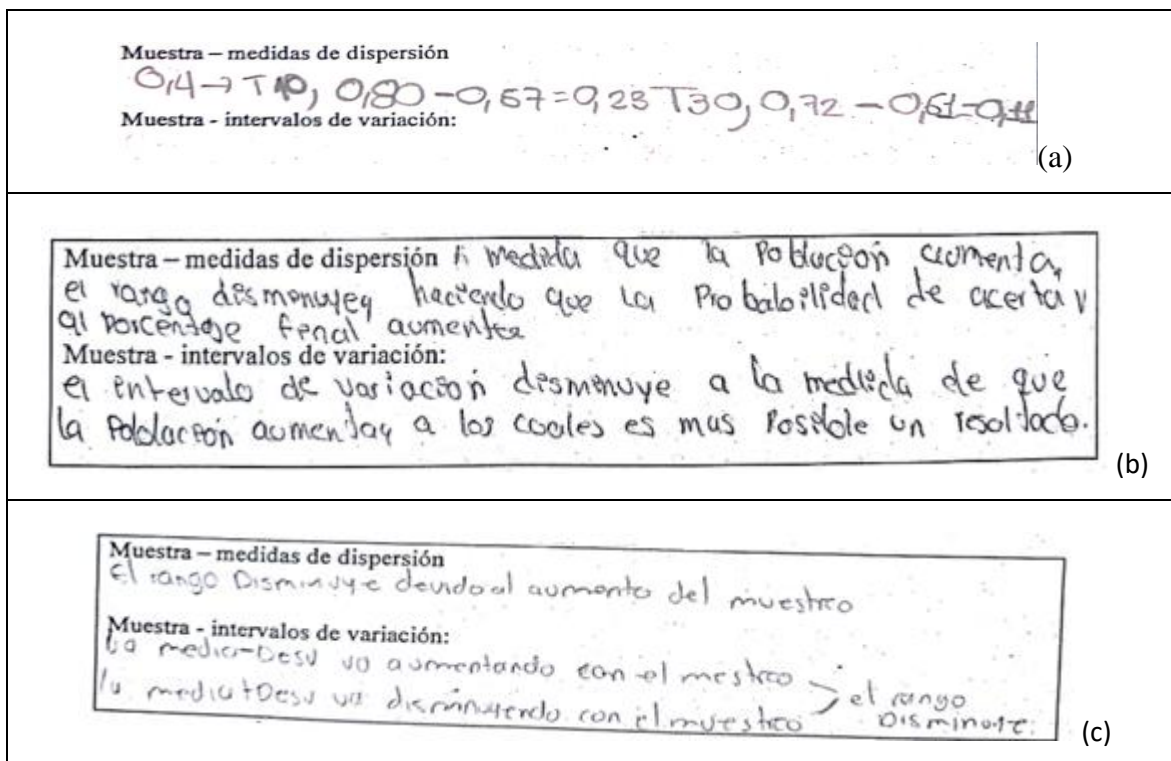
Comparación de las distribuciones para las proporciones de las 14 muestras de tamaños $n=10$, $n=30$ y $n=100$. Resultados obtenidos con los estudiantes del proyecto, febrero 2022



Los estudiantes manifestaron que no entendían la actividad que debían hacer. Por lo tanto, se les solicitó que leyeran las gráficas y hallaran diferencias entre sus distribuciones. Algunos estudiantes comentaron que “los valores más pequeños y los más grandes cambiaban”, refiriéndose a el rango; otros, alimentaron la idea diciendo que “cada vez, el error se hace más pequeño”, entendiendo por error el valor del rango. En ambas situaciones, la idea visual de la distribución apuntó a que el rango cambiaba según el tamaño de la muestra. Otro ejercicio que se propuso fue calcular el rango para cada distribución, siendo acertada esta decisión pues otros estudiantes comentaron que ahora sí habían entendido el literal. En este orden de ideas, los estudiantes reportaron en sus guías lo observado (Figura 21):

Figura 21.

Relaciones entre el tamaño del muestreo y las medidas de dispersión e intervalos de variación. Resultados de los estudiantes del proyecto, febrero 2022



Solo 8 estudiantes reportaron los cálculos explicados en el tablero (v.g. Figura 21.a). Los demás, concluyeron que mientras más grande sea la muestra tomada, el rango (medida de dispersión adoptada) disminuye, así como también lo hacen los intervalos de variación (v.g. Figura 21.b). Además, algunos grupos especificaron el comportamiento de los intervalos de variación, desde su cálculo (v.g. Figura 21.c).

Algunos estudiantes solicitaron ayuda porque no lograban establecer alguna relación entre los elementos pedidos. Por lo tanto, se les explicó de manera personal invitándolos a que centraran la atención en los valores mínimos y máximos que tomaba la variable (eje horizontal) y los intervalos de variación dados por el aplicativo en cada distribución. Los estudiantes analizaron lo comentado y finalmente, lograron reportar una respuesta.

Las conclusiones sobre las relaciones pedidas que expresaron los estudiantes fueron apropiadas desde una perspectiva disciplinar. Además, se evidencia una inclinación a usar

los datos como evidencia en la justificación de los argumentos sobre las relaciones entre el centro, su dispersión y el tamaño del muestreo.

Vale la pena aclarar que se decidió realizar el análisis de las medidas de dispersión alrededor del rango y no de la desviación estándar, ya que por motivos logísticos el tiempo destinado para el desarrollo de la sesión se redujo.

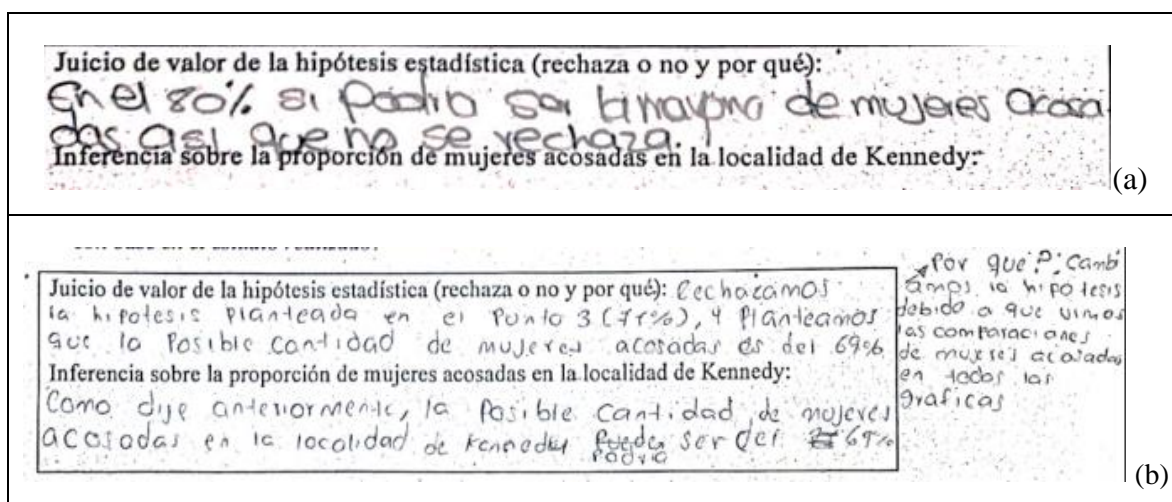
En este sentido, se leyó en voz alta el último punto de la guía, haciendo énfasis en la importancia de utilizar los elementos estadísticos estudiados para darle sentido a la respuesta.

5. Seleccione la proporción muestral que mejor pueda representar a la proporción poblacional y valide (argumentar si se rechaza o no) la última hipótesis planteada. ¿Entre qué valores podría estar la proporción poblacional de mujeres que consideran que ha habido casos de acoso callejero hacia las mujeres en la localidad de Kennedy con base en el estudio realizado?

A continuación, se ilustran algunas de las respuestas dadas por los estudiantes (Figura 22).

Figura 22.

Validaciones de las hipótesis estadísticas e inferencias sobre el parámetro estudiado. Resultados de estudiantes del proyecto, febrero 2022



Dos grupos realizaron la validación y conclusión sobre el parámetro en una misma oración (v.g. Figura 22.a). Los demás, lo reportaron en los espacios asignados para el ejercicio, valiéndose de las gráficas representativas de las distribuciones (v.g. Figura 22.b).

El estudio estadístico realizado con énfasis en la consideración del contexto tuvo un impacto positivo sobre el desarrollo del RII en los estudiantes, ya que se evidenció en ellos el pensamiento de incertidumbre alrededor de la situación problema y el sentido sobre los argumentos con los que validaron las hipótesis.

A la luz de los resultados obtenidos, se consideró para la siguiente sesión, hacer un mejor seguimiento sobre los estudiantes que aún presentaban dificultades con los procesos realizados, en particular con quienes insistían en el uso de justificaciones subjetivas.

Inferencia de diferencia de proporciones

Luego de haber realizado inferencias sobre una proporción, se trabajó con los estudiantes sobre diferencias de proporciones, como se propone en el siguiente enunciado:

- 1. Considere una muestra de tamaño 30 de mujeres que residen en la UPZ Patio Bonito y las muestras de tamaño relativa en las otras UPZ y halle las proporciones de las encuestadas que están de acuerdo con que en la zona ha habido casos de acoso. Realice una hipótesis de diferencia de proporciones entre (i) Patio Bonito- Américas, (ii) Kennedy Central-Patio Bonito. Ejemplos: la diferencia podría ser mayor que 0, puede no haber diferencia entre las proporciones, la diferencia de proporciones sería de -0,05.*

Los estudiantes registraron en la guía las proporciones correspondientes, realizaban a mano las diferencias solicitadas y proponían las hipótesis en el contexto de la situación (v.g. Figura 23); no obstante, solo el 36% de los 28 estudiantes realizó correctamente las diferencias. En consecuencia, para ayudar a los estudiantes que había no habían respondido correctamente, se realizaron en el tablero varias restas en el tablero, recordándoles que cuando el sustraendo es mayor que el minuendo, la diferencia es negativa.

Figura 23.

Hipótesis sobre diferencias de proporciones para muestras de tamaño $n=30$. Resultados de los estudiantes del proyecto, febrero 2022

UPZ	Proporciones
Américas	0,67
Kennedy central	0,60
Patio bonito	0,57

Hipótesis (i): $0,57 - 0,60 = -0,03$; Según la resta, la localidad más acosada entre P. bonito y Kennedy, es Kennedy

Hipótesis (ii): $0,57 - 0,67 = -0,1$; la localidad más acosada es Américas

la localidad menos acosada es P. bonito y la más acosada es Américas

A pesar de que los estudiantes respondían acertadamente a los ejemplos de restas dados, algunos insistían en registrar en las guías el valor absoluto de estas diferencias. Por lo cual, se decidió avanzar con el siguiente punto de la guía, ampliando esta explicación.

2. Suponga que para el caso de las diferencias proporcionales entre Kennedy Central y Patio Bonito se ha obtenido una diferencia negativa ¿Qué significa tener una diferencia menor que cero?

Más del 50 % de los estudiantes (18 de 28) comprendieron que la diferencia indicaría que el estimador estudiado en Patio Bonito era mayor que el de Kennedy Central, evidenciando un avance en el tratamiento del error epistémico mencionado en el literal anterior, así como también el desarrollo de la habilidad de contextualizar esta norma matemática (v.g. Figura 24). Los demás reportaron la diferencia con las proporciones registradas en sus tablas.

Figura 24.

Interpretación sobre la diferencia menor que cero. Resultados de los estudiantes del proyecto, febrero 2022

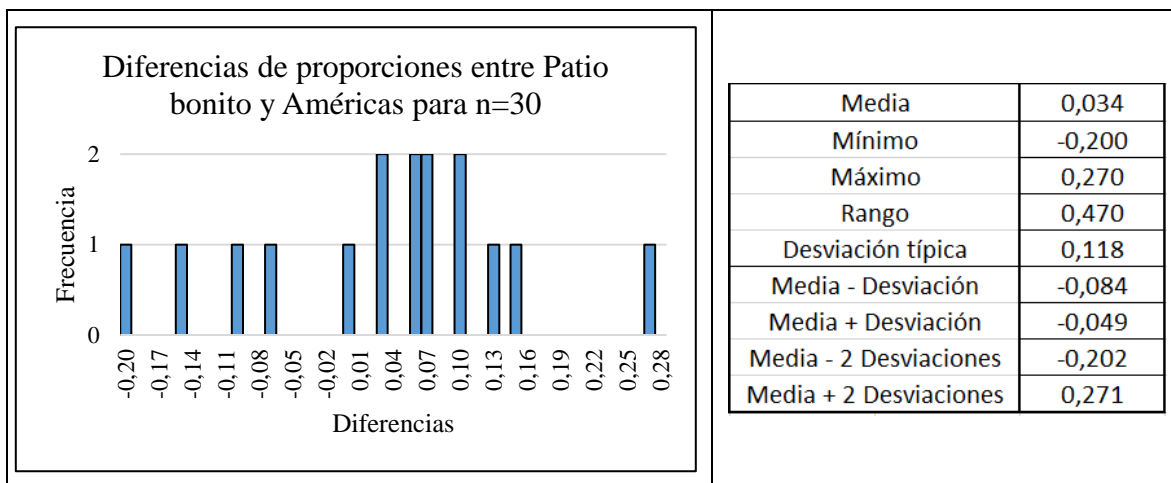
<p>Kenedi tiene un menor porcentaje de acoso que en patio Bonito respecto a la de Kennedy</p>	<p>Kennedy tiene un menor porcentaje de acoso que en Patio Bonito, respecto a la de Kennedy.</p>
---	--

Luego de observar las respuestas de todos los grupos, se le comentó al grupo que se estaba trabajando sobre el segundo punto de la guía y que la idea era contextualizar la idea de las diferencias estudiadas anteriormente en la situación problema.

Para orientar y contextualizar mejor la situación, se les pidió a los estudiantes que ingresaran las proporciones obtenidas en el ítem 1 al Excel proyectado. Allí, se les mostró en el gráfico estadístico la relación entre los valores de las proporciones y las diferencias, explicándoles una vez más el por qué algunas daban positivas, cero o negativas (v.g. Figura 25).

Figura 25.

Diferencia de proporciones Patio Bonito – Américas para una muestra de tamaño n=30. Resultados con los estudiantes del proyecto, febrero 2022



Después de la explicación los estudiantes procedieron a realizar el tercer ítem de la guía:

3. *Considere una muestra de tamaño 100 para la UPZ Patio Bonito y las muestras relativas de las demás UPZ y halle las diferencias de proporciones solicitadas en el ítem 1. ¿Cambiaría algunas de las hipótesis iniciales? Sí/ No ¿Por qué?*

Reportaron las nuevas proporciones (v.g. Figura 26), hallaron las nuevas diferencias y realizaron las correspondientes validaciones (v.g. Figura 27); solo dos grupos reportaron diferencias incorrectas. Para el caso de los estudiantes que no rechazaron sus hipótesis, evidenciaron que los porcentajes eran los mismos o muy similares a los obtenidos anteriormente, estando estos dentro de los posibles valores del parámetro. Mientras que los estudiantes que rechazaron su hipótesis reportaron las evidencias suficientes junto con las nuevas hipótesis. Una vez manifestaron haber registrado las respuestas en la guía, realizaron el juicio final sobre las diferencias entre los parámetros estudiados (v.g. Figura 28).

Figura 26.

Inferencia de diferencia de proporciones para una muestra de tamaño n=100. Resultados con los estudiantes del proyecto, febrero 2022

4. *Con base en los resultados obtenidos decida si acepta o rechaza alguna de las hipótesis planteadas sobre las diferencias de las proporciones. Justifique su respuesta.*

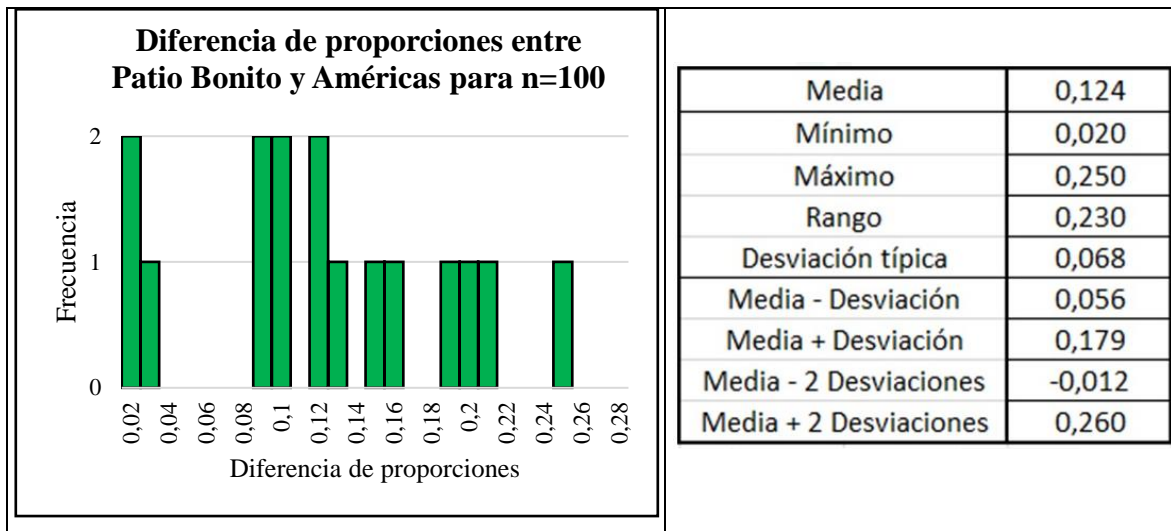


Figura 27.

Validación de hipótesis de diferencia de proporciones de tamaño n=100. Resultados de los estudiantes del proyecto, febrero 2022

$P-A=0.26$
 $K-P=-0.73$

cambio la hipótesis inicial 0,67% porque las localidades hoy diferentes personas por lo cual el porcentaje no puede ser 0 por lo que es más acertado nuestra nueva hipótesis

Si respondió Sí, escriba la nueva hipótesis

Hipótesis: Nuestra nueva hipótesis es cambiar porque es porcentaje = 0.26 y 0.73 es más acertado por la desigualdad de localidades

Figura 28.

Argumentación sobre el juicio final dado a las hipótesis planteadas para diferencia de proporciones. Resultados de los estudiantes del proyecto, febrero 2022

Juicio sobre la hipótesis 1: Patio Bonito - Américas
 Mantengo mi hipótesis de 0,13 debido a que tomo en cuenta la grafica y me siento que esta muy bien en proporción

Juicio sobre la hipótesis 2: Kennedy Central - Patio Bonito
 Mantengo nuestra hipótesis de -0,17 tambien basandome en la grafica y sus valores

Los estudiantes se basaron en las gráficas obtenidas para evidenciar las decisiones tomadas sobre las hipótesis propuestas. Se evidenció una disminución en la frecuencia del error epistémico asociado a la diferencia entre proporciones marcado al iniciar la sesión.

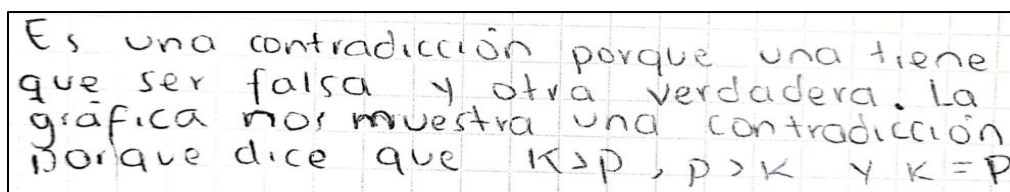
Para finalizar la sesión, se les propuso a los estudiantes estudiar el siguiente caso:

5. *Suponga que el intervalo entre diferencias de proporciones entre dos UPZ es de $(-0.1, 0.1)$. Explique qué se puede concluir sobre la diferencia de proporciones en la población entre las dos UPZ para este caso.*

Algunos no tuvieron problemas en identificar que el “0” pertenecía a este intervalo (el cual era una de las intencionalidades); por lo tanto, se les representó mediante la recta numérica este intervalo, recordándoles que éste se conforma por los elementos que están entre sus extremos. De esta manera, la gran mayoría de los estudiantes manifestaron verbalmente que el “0” debería estar en ese intervalo. Luego, en los grupos de trabajo establecieron discusiones sobre qué significaría reportar en el estudio el intervalo, pues se podría interpretar que en la población hay mayor, igual y menor diferencia entre las UPZ, siendo este hecho contradictorio (v.g. Figura 29). Además, fue fácil evidenciar por la mayoría que el “0” significaba no haber diferencia alguna entre las proporciones, ya que este evento ocurrió en algunos muestreos realizados (v.g. Figura 30).

Figura 29.

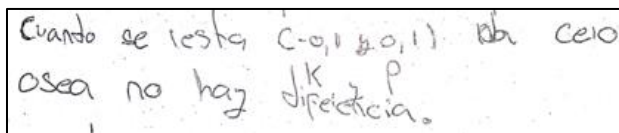
Argumentación de la contradicción encontrada por los estudiantes para el intervalo $(-0.1, 0.1)$. Resultados de los estudiantes, febrero 2022



Es una contradicción porque una tiene que ser falsa y otra verdadera. La gráfica nos muestra una contradicción porque dice que $k > p$, $p > k$ y $k = p$

Figura 30.

Ejemplo interpretación de la existencia del “0” en el intervalo $(-0.1, 0.1)$. Resultados de los estudiantes, febrero 2022



Quando se resta (-0,1 y 0,1) da cero
o sea no hay diferencia.

En cada grupo de trabajo, se revisaron las respuestas dadas por los estudiantes en las guías, evidenciando respuestas similares a las expuestas en el análisis de esta sesión lo cual evidencia una apropiada comprensión acerca de la interpretación de los intervalos de variación y la formulación y validación de hipótesis sobre inferencias de diferencias de proporciones. En este sentido, se les agradeció la participación a los estudiantes y se les comentó que la próxima semana tendrían la última sesión.

Guía tres

Inferencia sobre las medias poblacionales con una muestra de tamaño 100

En la última sesión se realizaron inferencias sobre las medias poblacionales correspondientes a la edad de las mujeres de la UPZ Castilla que consideran haber sufrido de acoso callejero bajo el siguiente enunciado:

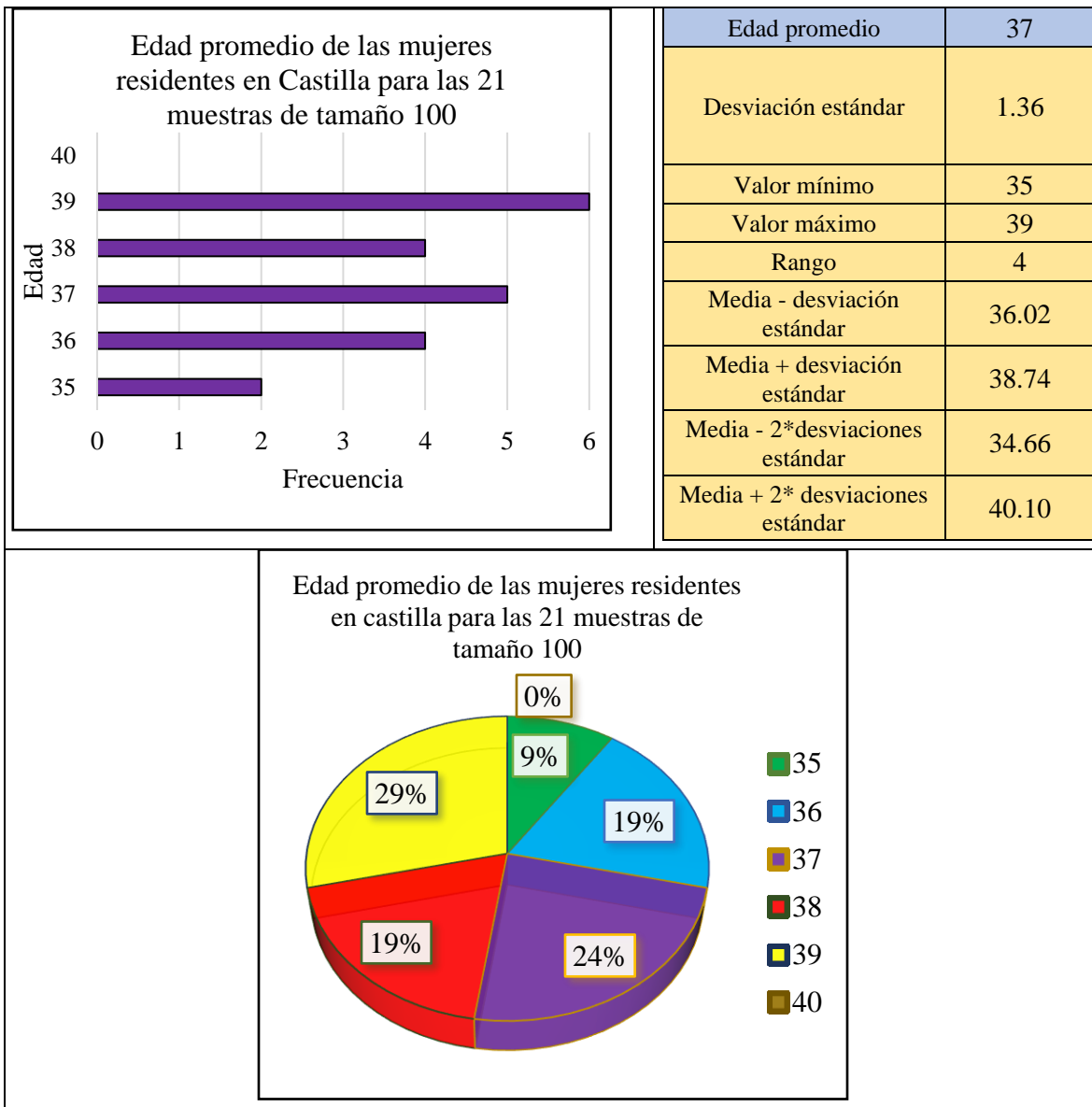
Considere como hipótesis la siguiente afirmación:

Las mujeres que habitan en la UPZ Castilla que dicen que en la zona se han presentado casos de acosos hacia las mujeres, tienen una edad promedio mayor o igual que 42 años.

Los estudiantes realizaron el muestreo de tamaño 100 con la ayuda del aplicativo, reportaron las medias que obtuvieron e ingresaron los demás resultados de sus compañeros, los cuales estaban proyectados en el tablero. Algunos grupos hicieron doble muestro para obtener así más datos que aportaran al análisis. Además, registraron en las guías la media de los datos obtenidos del muestro realizado, el promedio de estas medias, las medidas de dispersión y los intervalos de variación, realizando así los primeros dos puntos de la guía. A continuación, se ilustran las representaciones gráficas obtenidas (Figura 31).

Figura 31.

Representaciones gráficas del muestreo conjunto y tabla de resumen para la inferencia de medias. Resultados obtenidos con los estudiantes del proyecto, marzo 2022



Luego, continuaron con el tercer enunciado:

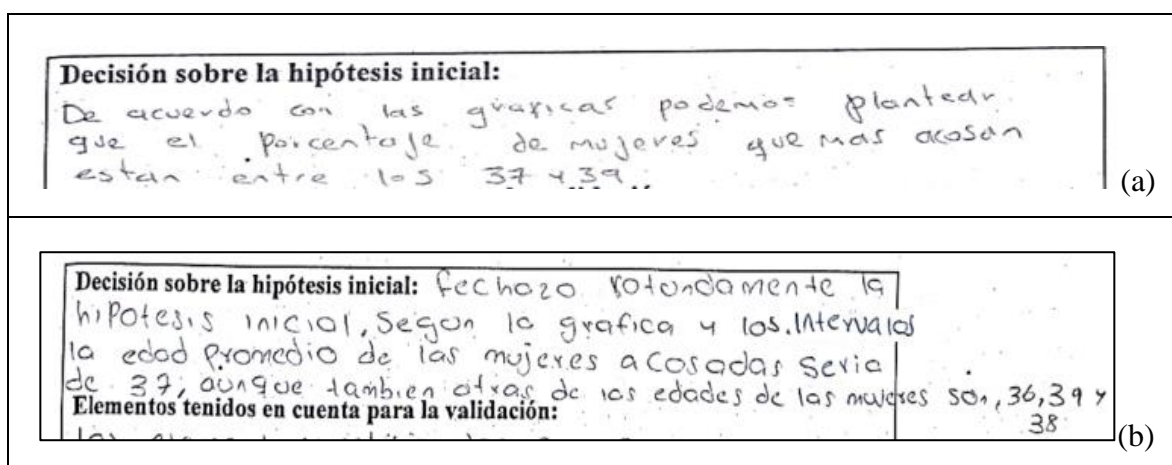
3. *Argumente si rechaza o no la hipótesis planteada y comente en que elementos estudiados (frecuencias absolutas o porcentuales de las gráficas, medidas de dispersión, medidas de centralidad, intervalos de variación, etc.) se basó para dar esta validación.*

Tres grupos de los catorce reportaron como porcentajes las edades enunciadas (v.g. Figura 32.a); lo anterior suele ser un error común, ya que los estudiantes estaban acostumbrados a realizar inferencias sobre proporciones, dificultando así la nueva contextualización. A ellos se les recordó que la variable de inferencia con la que están trabajando es la edad de las

mujeres, y no si es o no acosada; esta diferencia lleva por supuesto, a distinguir que ahora se hable de promedios de edades y no de proporción de acosos. Para el caso de los demás estudiantes, rechazaron la hipótesis inicial, reportando como evidencias las representaciones gráficas y los intervalos de variación obtenidos del muestreo de todo el grupo (v.g. Figura 32.b).

Figura 32.

Decisiones sobre la hipótesis sobre la media estudiada con base en el muestro de tamaño $n=100$. Resultados de los estudiantes del proyecto, marzo 2022



Se evidenció que los estudiantes tuvieron en cuenta para su decisión todos los valores posibles ilustrados y no solo el valor numérico de la frecuencia de la moda, como lo hacían en las guías anteriores. Después de socializar la decisión sobre la hipótesis, a los estudiantes se les invitó a enunciar una conclusión, tal como se presenta en el siguiente enunciado.

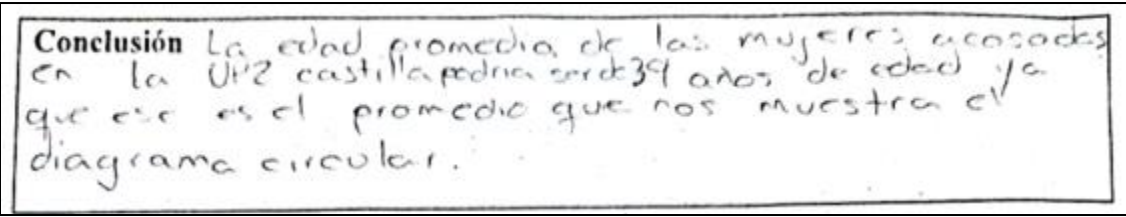
- 4. Realice una conclusión para el parámetro estudiado, las edades promedio de las mujeres encuestadas que residen en la UPZ Castilla que están de acuerdo con el hecho de que en la zona se han presentado escenas de acoso hacia las mujeres, con base en los resultados obtenidos en el estudio de la muestra. Utilice la información suministrada por los intervalos de variación y por las gráficas ilustradas para justificar dicha conclusión.**

A pesar de haber considerado varias edades posibles en el ítem anterior, varios estudiantes concluyeron identificando únicamente la moda, para de esta manera atribuir un solo valor posible (v.g. Figura 33). En este sentido se observa una falta de consideración de la observación de la variación alrededor del valor de la moda por parte de los estudiantes. Para

subsanan este error, se les comentó a los estudiantes que para la conclusión deberían considerar la variación reportada en la decisión, ya que es con base a ese argumento y a los objetivos que se plantearon con la hipótesis que se construye la conclusión, o idea sobre el comportamiento de la edad de la población.

Figura 33.

Conclusión sobre la hipótesis inicial de la edad promedio de las mujeres. Resultados de los estudiantes del proyecto, marzo 2022



Conclusión La edad promedio de las mujeres acosadas en la UP2 castilla podría ser de 39 años de edad ya que ese es el promedio que nos muestra el diagrama circular.

Debido a que en este momento algunos estudiantes habían realizado rápidamente las conclusiones, tuvieron autonomía para continuar con el desarrollo del resto de la guía, realizando las preguntas necesarias en el proceso. De esta manera, se tomó la decisión de que cada grupo continuara su trabajo con las preguntas propuestas en la guía. Así, ellos siguieron su trabajo con el siguiente enunciado.

5. *Con base en los resultados obtenidos discuta sobre:*
 - (i) *¿De qué edad(es) promedio sería menos probable encontrar una mujer que esté de acuerdo con que en el sector las mujeres son víctimas de acoso, y*
 - (ii) *¿Qué porcentaje de estas mujeres representan?*

Para cada caso, mencione que tuvo en cuenta para dar su respuesta.

Hubo estudiantes que estaban realizando los promedios a mano; sin embargo, comentaron que el ejercicio les parecía muy largo para desarrollarse. En este sentido, se les invitó a observar y leer el diagrama circular; así, ellos identificaron que este representaba los porcentajes de las edades reportadas en el muestreo conjunto. Otros, argumentaron con base en el intervalo de variación y reportaron los valores leídos en el diagrama circular (v.g. Figura 34).

Figura 34.

Argumento sobre la menor probabilidad de edad. Resultados de los estudiantes del proyecto, marzo 2022

(i) La edad de mujeres menos probable que han sido acosadas es de 40 años.

(ii) Según el diagrama circular las mujeres de 35 años tienen un 9%, las de 36 años es de 19%, las de 37 años tienen un 24%, las de 38 tienen un 19%, las de 39 años tienen un 29% y las de 40 años tienen un 0%.

En el contexto que se venía trabajando (acoso), la adquisición del lenguaje probabilístico se continuó evidenciando mediante expresiones como “menos probable”, “podría ser”, “sería”. Además, utilizaron correctamente los datos como evidencia para darle fuerza a sus argumentos, tras haber sistematizado la información dada al inicio del estudio realizado.

Luego, cada grupo resolvió el siguiente ítem de la guía:

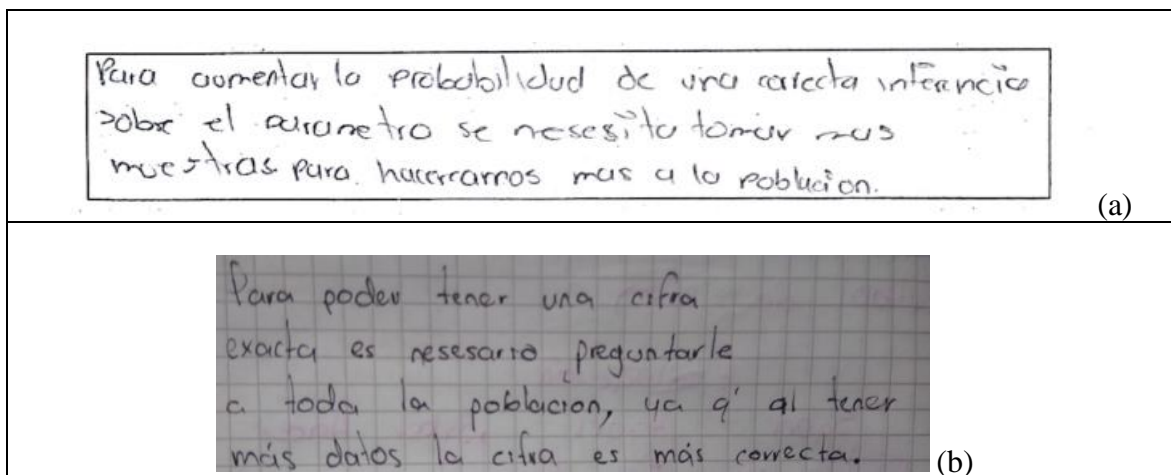
6. *Considere lo aprendido en las guías trabajadas y responda ¿Cómo se puede mejorar la calidad del estimador, considerando el tamaño de la muestra?*

Hubo estudiantes que preguntaron sobre los conceptos de estimador y parámetro, pues no se acordaban de su significado. Por lo tanto, se les explicó que el estimador era la herramienta (estadístico) que permitía estimar el valor numérico de la característica que se estudiaba en la muestra, siendo en este caso la media de las edades de las mujeres; mientras que el parámetro hacía referencia a el valor de esta característica en la población.

La mayoría de los estudiantes comprendieron que, a mayor tamaño de muestra, menor sería el rango de la distribución (lo cual muchos interpretaban como el error) (v.g. Figura 35.a). Sin embargo, fueron pocos los que precisaron este hecho argumentando que, si consideraban a toda la población en el estudio la inferencia sería “más exacta” (v.g. Figura 35.b).

Figura 35.

Argumentos sobre la obtención de una mejor inferencia. Resultados de los estudiantes del proyecto, marzo 2022



Dada la autonomía con la que estaban trabajando los grupos, se realizaron comentarios en cada uno de ellos sobre los argumentos que realizaron. Para el caso de aquellos que no daban respuesta o las daban de manera muy imprecisa, se les invitó a comparar los gráficos estadísticos obtenidos en la comparación de las proporciones de mujeres acosadas para muestras de tamaño 10, 30 y 100 y se les preguntó “¿cuántos datos deberían tomar para que el rango fuera de 0?”, ante lo cual, tras discutir un buen tiempo, respondieron que “deberían tomar a todas las mujeres”. Se les aclaró que eso significaba tomar a toda la población, y así se estaría hablando del parámetro y no del estimador.

Una vez que los grupos comprendieron esta precisión, siguieron al último punto de la guía.

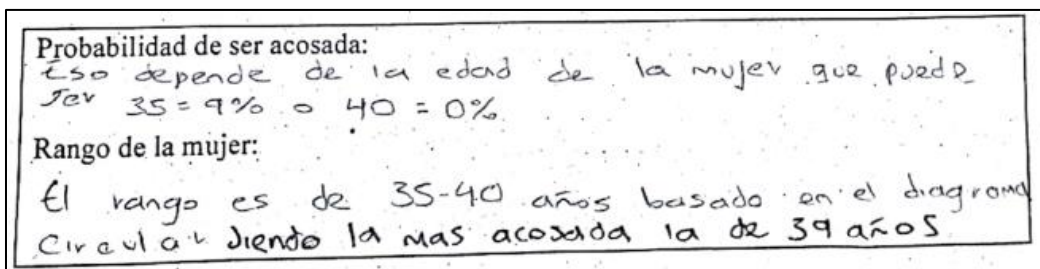
7. Suponga que una mujer que va de compras cerca al barrio Portal de las Américas, el cual está ubicado en la UPZ # 46: Castilla. Si las distribuciones de las edades de las encuestadas y de las proporciones en la localidad de Kennedy se comportan similar es a los de la muestra de tamaño 100 estudiadas (son significativas). ¿Cuál podría ser la probabilidad de que la mujer sea acosada? ¿Cuál es el rango de edad más probable de esta persona?

Los estudiantes dieron diversidad de respuestas, argumentando que la probabilidad de ser acosada dependía de la edad de las mujeres. Además, reportaron la edad relacionada a dicha probabilidad, o en su defecto, el rango de edad obtenido en el análisis de datos (v.g. Figura

36). A pesar de ello, fueron conscientes de que podían utilizar la información obtenida en el estudio anterior, evidenciando así un correcto razonamiento.

Figura 36.

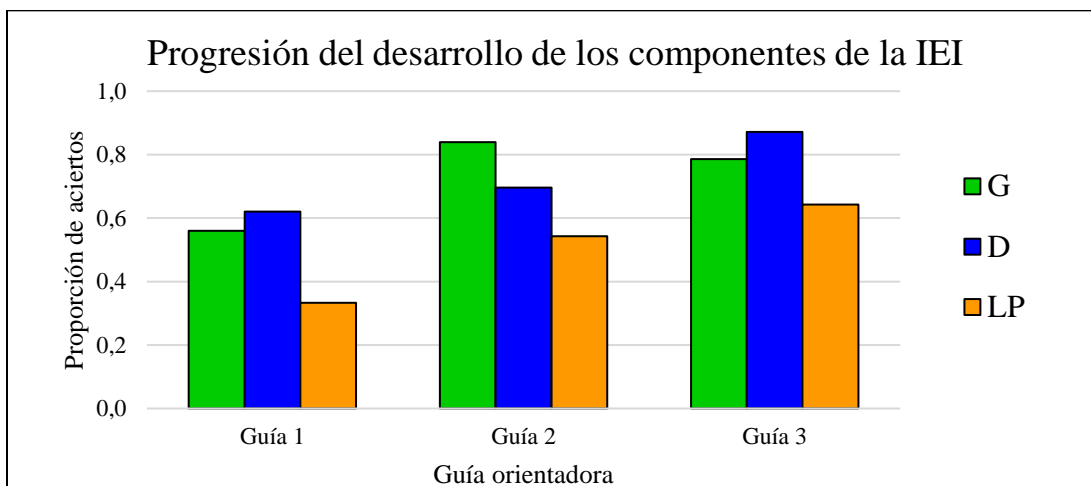
Ejemplo de respuestas para el ítem 7: mujeres de compras en Castilla. Resultados de los estudiantes del proyecto, marzo 2022



Por un lado, se cumplió la intencionalidad del ítem ya que los estudiantes hicieron uso de los resultados obtenidos en el estudio anterior. Por otro lado, la diversidad de respuestas obedece a la manera en cómo está expresada la pregunta, ya que como se evidenció, puede dar lugar a varias interpretaciones.

Figura 37.

Progresión del desarrollo de los componentes de la IEI. Resultados obtenidos de los estudiantes del proyecto, marzo 2022



Nota. G: Generalización; D: Uso de datos como evidencia; LP: Lenguaje probabilístico

Finalmente, se muestra la progresión objetiva de los componentes de la IEI: G, D y LP evidenciado en las respuestas de los estudiantes a lo largo del proyecto (Figura 37). Por un lado, en la última guía se observó una leve desmejora en las habilidades asociadas a la generalización, debido a que los estudiantes estaban desarticulando el contexto de las afirmaciones o validaciones. Por otro lado, aquellas asociadas al uso de datos como evidencia y al lenguaje probabilístico presentaron una mejora continua.

Conclusiones

En este trabajo se logró la elaboración e implementación de una secuencia de actividades inspiradas en una situación de interés de los estudiantes, las cuales tuvieron como propósito general ejemplificar los actos de enseñanza y aprendizaje implicados en el desarrollo del razonamiento inferencial informal de los estudiantes y en las habilidades que enuncian los documentos curriculares para el grado décimo en el marco de los sistemas de datos y del pensamiento aleatorio.

La manera en cómo se orientó el desarrollo de las guías orientadoras dio cuenta del progreso de las habilidades relacionadas a la generalización más allá de los datos (planteamiento y validación de hipótesis, conclusiones); al uso de los datos como evidencia en el proceso de argumentar sobre la justificación o refutación de las hipótesis; y a la apropiación del lenguaje probabilístico implicado en los procesos de argumentación, interpretación y validación. Además, permitió identificar dificultades y obstáculos que presentaron los estudiantes en la adquisición del RII (diferenciar parámetro de estimador, población de muestra), así como algunos errores consulados en la literatura (desarticulación del lenguaje probabilístico del contexto, juicios y validaciones subjetivos). Por último, se logró realizar un ejercicio de análisis alrededor de la puesta en marcha, al hacer un contraste entre los resultados y las intencionalidades enunciadas, a la luz de la literatura consultada.

En ese sentido, los componentes considerados en los marcos conceptuales y metodológicos fueron suficientes para el diseño de las actividades que conformaron las tareas propuestas: prueba diagnóstica y secuencia de actividades. En particular, aquellos relacionados con la distribución, dispersión, hipótesis estadísticas y cada referente de la educación estadística. Gracias a la apropiación de conocimientos procedentes de la literatura consultada, se realizaron acciones didácticas razonables tanto para la puesta en marcha de la secuencia de actividades, como en la gestión misma de las actividades. Asimismo, se identificaron los componentes y las dificultades, obstáculos y errores asociados al aprendizaje de la IEI.

Respecto a los resultados del trabajo con los estudiantes, se consideran las siguientes observaciones: La progresión de las habilidades implicadas en la estadística inferencial desarrollada en la medida en que ellos participaban de las experiencias propuestas, en consonancia con lo planteado por Sánchez y Ruiz (2017).

Los estudiantes reconocen la importancia de la presencia de representantes de cada estrato para poder inferir sobre el parámetro (proporción poblacional), a partir del estimador obtenido (proporción muestral), concepción que obedece a lo planteado en la literatura (v.g. Mendenhall et al., 2010; Levine et al., 2006, entre otros). Asimismo, son capaces de relacionar los muestreos de diferentes tamaños con el rango y los intervalos de variación que caracterizan la distribución y de planear estrategias para minimizar el rango a tal punto de reconocer diferencias entre muestra y población y estimador y parámetro. Se recomienda trabajar alrededor de esta medida de dispersión en las primeras inferencias, ya que según Anderson et al. (2010), el rango es la medida de dispersión más sencilla con la que se puede argumentar la variación.

Es común que en las primeras experiencias los estudiantes aludan a la descripción de la población, mediante expresiones que refieran a su caracterización (exactamente, el porcentaje de la población es, la media es exactamente), e inclusive, opten por el valor de la moda de una distribución de muestreos, considerándolo como el único estimador posible. Ante estos escenarios, es urgente realizar las respectivas aclaraciones, ya que Batanero et al. (1993; citado por Sánchez y Ruiz, 2017) marca que, como consecuencia de la transición del razonamiento descriptivo al inferencial, los estudiantes presentan dificultades para identificar el estimador del parámetro y de interpretar correctamente las representaciones gráficas de las distribuciones.

Los estudiantes suelen desarticular el contexto o la interpretación del análisis de datos del lenguaje probabilístico, el cual es un error común según lo consultado en García y Sánchez (2014). La insistencia en el correcto uso de este lenguaje y las precisiones anunciadas sobre la inferencia desde una muestra, son acciones claves y urgentes por realizar, puesto que, según Garfield & Ben – Zvi, (2011; citado por García, 2014), si no se trata con premura, los estudiantes no desarrollarán correctamente el RII en especial se les dificultará presentar validaciones adecuadas.

También se encontró que ellos obvian usar los datos como evidencia y deciden argumentar alrededor de sus creencias, desembocando en lo que según Zieffler et al. (2008) denominan justificaciones subjetivas. En consecuencia, desarticulan el planteamiento de hipótesis y sus juicios de valor del contexto de la situación, haciendo que el ejercicio carezca de sentido común, en consonancia con lo que comenta Dewey sobre la importancia del contexto en la realización de inferencias estadísticas (1910; citado por Makar et al. 2009). Por lo tanto, son acciones que se deben considerar y dar prioridad en la orientación hacia el desarrollo de argumentaciones fundamentadas, propias de una persona estadísticamente culta (Contreras y Molina, 2013).

Una buena estrategia para abordar esta desarticulación entre el planteamiento de hipótesis y los respectivos juicios de valor de la situación es el uso de software estadístico ya que favoreció que los estudiantes centraran la atención sobre los gráficos estadísticos y las tablas resumen, indicando que la visualización juega un papel importante en la interpretación de la información estadística (Batanero, 2001). Por lo tanto, se recomienda abordar los elementos estadísticos que sirven de datos como evidencia desde estas representaciones, puesto que se mostró que de esta manera los estudiantes adquieren una visión consciente del contexto estudiado y desarrollan validaciones coherentes, adquiriendo una posición crítica y cuestionadora sobre la información estadística, en concordancia con lo sugerido por Contreras y Molina (2013).

Como consecuencia de estas acciones didácticas, los estudiantes lograron valerse de las inferencias realizadas para comunicar el posible comportamiento del fenómeno social estudiado (acoso), integrando conocimientos adquiridos en la estadística descriptiva (dispersión, centro, moda, rango), lo cual es muestra de la adquisición epistémica de la inferencia informal, en el sentido en el que lo refieren Zieffler et al. (2008).

En consideración a lo anterior, se aprecia que realizar inferencias en torno a un contexto de interés para los estudiantes permite potenciar el uso de datos como evidencias, la formulación y juicio de hipótesis, la lectura de y entre datos y la comunicación con el lenguaje probabilístico mediante expresiones como “es más acertado”. Es decir, se evidenció un desarrollo en las habilidades asociadas al RII, propuestas por Pfankuch (2006, citado en Sánchez y Ruiz, 2017).

Cabe mencionar que el acompañamiento continuo hacia los estudiantes en el avance de las habilidades relacionadas al RII y a la IEI es vital, puesto que pueden desarticular el contexto del lenguaje probabilístico, o de los conocimientos del saber estadístico aprendido al enfrentarse a nuevas situaciones.

Finalmente, se reconoce que a lo largo de este trabajo se desarrollaron aptitudes investigativas (identificación de una problemática, diseño de un plan de acción con base en referentes, obtención y análisis de datos, reporte de conclusiones), didácticas (identificación y acción alrededor de dificultades, obstáculos y errores, instrumentalización de Software para el diseño y aplicación de una secuencia de actividades, orientación en la construcción de conocimientos), y socioculturales (comunicación asertiva verbal y escrita, empatía, persuasión, promoción de valores) las cuales son fundamentales para realizar el ejercicio docente.

Consideraciones generales

Se reconoce que un estudio en el que se relacionen el muestreo aleatorio simple y el muestreo por estratificación debería realizarse en la introducción a las hipótesis estadísticas, ya que el reconocimiento y elección del muestreo puede fortalecer las habilidades de argumentación y promover en el estudiante una actitud crítica y cuestionadora sobre la información estadística.

En ese sentido, se sugiere que se le dedique una de las sesiones de clase al estudio de estos muestreos la cual debería preceder de la introducción del concepto de hipótesis estadística. De igual manera, el estudio de la desviación estándar como parámetro o elemento a considerar en el análisis de datos podría tener lugar en el estudio de la inferencia de medias, e inclusive desde el estudio de la inferencia de proporciones, según el nivel de desarrollo del pensamiento aleatorio y sistemas de datos que caracterice a los estudiantes.

Entre las acciones de un posible plan de mejoramiento para compensar la “leve desmejora” que se enuncia al finalizar el apartado de resultados podrían estar:

- ✓ Formular preguntas orientadoras que relacionen el contexto de la situación problema con las hipótesis estadísticas propuestas por el guía (profesor) o por los estudiantes.
Por ejemplo: ¿será cierto que el porcentaje de mujeres acosadas en Patiobonito será

mayor que el de las mujeres acosadas en Américas? ¿Por qué?, ¿están de acuerdo de que las mujeres de Castilla con más de 35 años no consideran ser víctimas de acoso? ¿En qué se basan para dar su respuesta?

- ✓ Rescatar los argumentos dados por cada uno de los grupos en el tablero y pedir a los estudiantes que validen dichos, recordándoles cuando se considere necesario el contexto de la situación.
- ✓ Guiar a los estudiantes en la redacción de un informe estadístico en el que reporten la situación problema, la población de estudio, las hipótesis estadísticas y los estimadores de estudio, los muestreos realizados, el análisis de datos y su interpretación, la validación de las hipótesis planteadas apoyándose de los datos como evidencia y del lenguaje.

Referencias

- Anderson , D., Sweeney, D., & Williams, T. (2010). *Estadística para administración y economía 10a. edición*. México D.F.: Cengage Learning.
- Andrade, L., Fernández, F., y Méndez, M. (2021). Exploración de la noción de distribución desde la variabilidad. *Tecné, Episteme y Didaxis: TED*, 4, 73-90.
- Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. (D. Springer, Ed.) *The challenge of developing statistical literacy, reasoning and thinking*, 147-168.
- Batanero, C. (2001). *Didáctica de la Estadística*. Granada: Grupo de Investigación en Educación Matemática. Departamento de Didáctica de la Matemática. Universidad de Granada.
- Colegio Isabel II I.E.D. (2019). *Proyecto educativo institucional*. Bogotá D.C. Colombia.
- Contreras, J. M., y Molina, E. (2019). Elementos clave de la cultura estadística en el análisis de la información basada en datos. En J. M. Contreras, M. M. Gea, M. M. López-Martín y E. Molina-Portillo (Eds.). *Actas del Tercer Congreso Internacional Virtual de Educación Estadística*. Disponible en <http://www.ugr.es/local/fqm126/civeest.html>.
- Contreras, J. M., Cañadas, G. R., Gea, M. M., y Arteaga, P. (2013). Inferencias estadísticas informales en estudiantes mexicanos. *Actas de las Jornadas Virtuales en Didáctica*, 343-357. Granada, Departamento de Didáctica de la Matemática de la Universidad de Granada.
- Fajardo, D. M. (2014). *Unidad didáctica para la enseñanza de la inferencia estadística en contextos biológicos*. Tesis de maestría. Universidad Nacional de Colombia. Bogotá D.C.
- Flores, A., y Pinto, J. (2017). Características de la enseñanza de la estadística por proyectos. En Serna, Luis Arturo (Ed.), *Acta Latinoamericana de Matemática Educativa*, 263-271. México, DF: Comité Latinoamericano de Matemática Educativa.

- García, V., y Sánchez, E. (2014). Razonamiento inferencial informal: el caso de la prueba de significación con estudiantes de bachillerato. *Investigación en Educación Matemática XVIII*, 345-354.
- I.E.D Nuevo Chile (2018). *Proyecto educativo institucional*. Bogotá D.C. Colombia.
- Levine, D., Krehbiel, T., & Berenson, M. (2006). *Estadística para administración: Cuarta edición*. México: Pearson Educación.
- Makar, K., & Rubin, A. (Mayo de 2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, 8(1), 82-105.
- MEN. (2006). *Estándares Básicos de Competencias en Lenguaje, Matemáticas, Ciencias y Ciudadanas*. Bogotá D.C.: Ministerio de Educación Nacional.
- MEN. (2016). *Derechos Básicos de Aprendizaje V.2*. Bogotá D.C.: Ministerio de Educación Nacional.
- Mendenhall, W., Beaver, R., & Beaver, B. (2010). *Introducción a la probabilidad y estadística. 13a. Edición*. México D.F.: Cengage Learning.
- Reading, C., & Shaughnessy, J. (2004). Reasoning about variation. *The challenge of developing statistical literacy, reasoning and thinking*, 201-226.
- Sánchez, N., y Ruíz, B. (2017). La inferencia informal en la enseñanza de la estadística. Una propuesta por medio del estudio de clases. *Avances en Educación Matemática. El profesor investigador.*, 6, 117-133.
- Wild, C. (2006). The concept of distribution. *Statistics Education Research Journal*, 5(2), 10-26.
- Zieffler, A., Garfield, J., Delmas, R., & Reading, C. (Noviembre de 2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 40-58.

Anexos

1. Prueba diagnóstica

Nombre: _____ Fecha: _____

1) Relaciona cada tabla de frecuencias con su respectivo diagrama de barras.



1

A

Fruta para jugo	Frecuencia
Mango	5
Mora	1
Limón	9



2

B

Lugar para vacaciones	Frecuencia
Cartagena	6
Amazonas	8
Villavicencio	4



3

C

Deporte para clase	Frecuencia
Baloncesto	7
Béisbol	10
Fútbol	3



Diagrama de barras	Tabla de frecuencias
	A
	B
	C

2) El siguiente pictograma representa el deporte favorito de un grupo de alumnos.



Considerando esta información, responde:

Pregunta	Procedimiento
i) ¿Cuántas personas practican Básquet? R//	
ii) ¿Cuál es la diferencia entre las personas que practican Fútbol respecto a las que practican Tenis? R//	
iii) ¿Cuántos alumnos se encuestaron en total? R//	

3) Esta es la tabla de datos que obtuvo Camila, al preguntar a 10 de sus compañeros por el tamaño de sus mascotas. Suma los elementos de cada categoría y reporta estos resultados en la fila “**Total**”

Encuestados	Grande	Mediano	Pequeño
Pedro		X	
Angélica		X	
Miguel	X		
Camilo		X	
Felipe			X
Marcela	X		
Diana	X		
Luisa		X	
Manuel			X

Ricardo		X	
Total			

4) A continuación se presenta el diagrama tabular que propone Camila, ordénala de mayor a menor según la frecuencia de las categorías y justifica si es verdad o falsa la afirmación “Las mujeres prefieren los perros pequeños, mientras que los hombres prefieren los perros grandes”.

Tabla de frecuencias			
Personas \ Categorías			
Hombres			
Mujeres			
Total			

La afirmación es _____ porque

5) Considera el siguiente diagrama, construye un diagrama de torta y responde ¿Cuál es el porcentaje de deportes que no requieren de una pelota?

R// _____

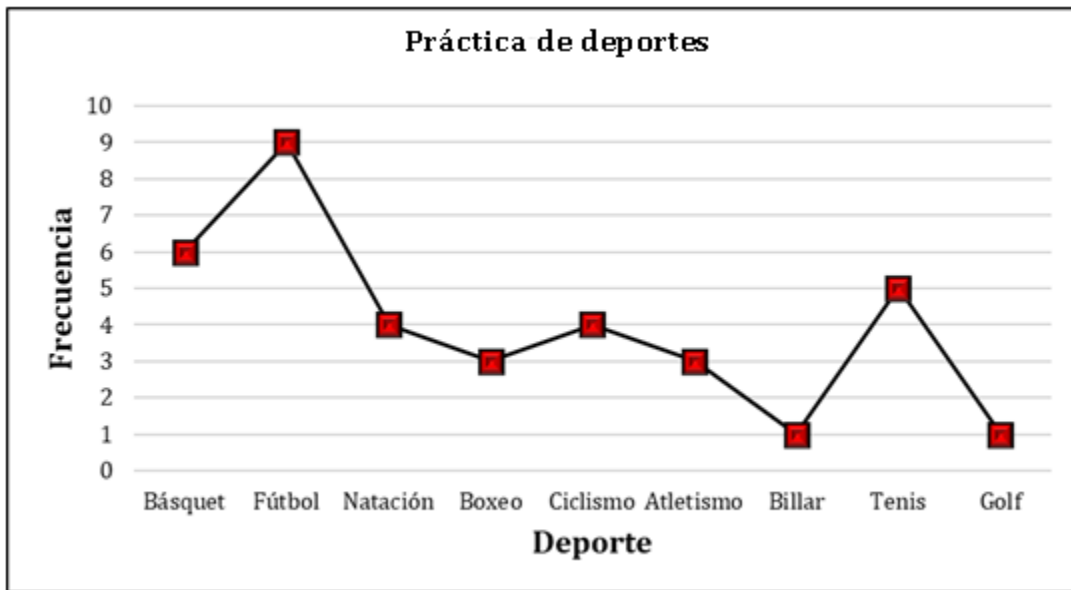
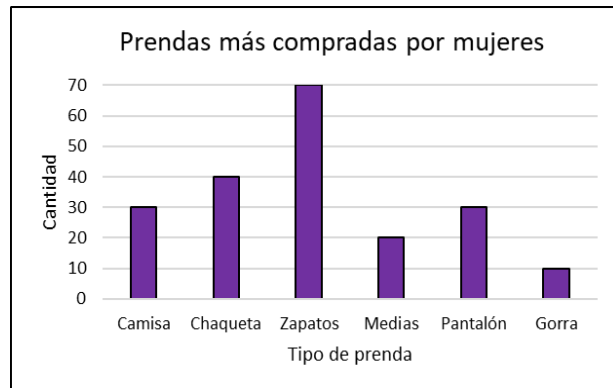
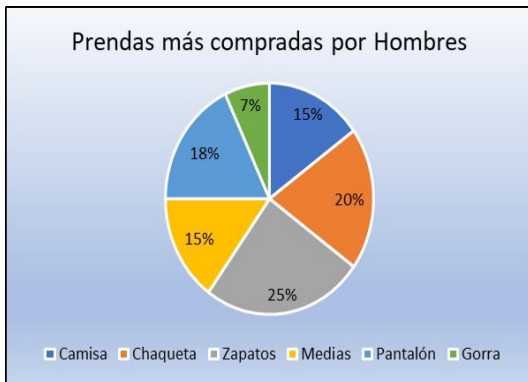


Diagrama de torta	Procedimiento

Resuelve los puntos 6 al 11 considerando el siguiente enunciado:

Se tomó una muestra de igual cantidad de mujeres y de hombres que frecuentan comprar la ropa en los Outlet de las Américas y se les preguntó sobre el tipo de prenda que más suelen comprar. A continuación, se reportan los resultados:



Con base a la información anterior responde las siguientes preguntas:

- 6) ¿Cuál es la moda entre las prendas de las mujeres? R// _____
- 7) ¿Cuántos hombres prefieren comprar chaquetas en lugar de comprar medias?
R// _____
- 8) ¿Cuál fue la cantidad total de personas encuestadas? R// _____

Procedimiento:

2. Guía uno

Nombres y apellidos:

Curso:

Parte 1: Inferencia de proporciones con una muestra de tamaño 10 (plantilla tamaño 10)

1. *Un estudio hecho por la Secretaría de la Mujer reveló alarmantes cifras sobre la consideración de acoso callejero a mujeres en Bogotá. Plantee una afirmación acerca de la proporción (o porcentaje) de mujeres de la localidad de Kennedy que usted piensa que consideran que en la zona ha habido casos de acoso callejero hacia las mujeres. Ejemplos: el 0.4 de las mujeres considerarían que en la zona ha habido casos de acoso callejero hacia las mujeres, más del 40% considerarían que en la zona ha habido casos de acoso callejero hacia las mujeres.*

Hipótesis:

Para responder las preguntas que siguen, utilice el archivo de Excel que fue preparado para tomar muestra de tamaño 10 en un grupo de 1372 mujeres de la localidad de Kennedy.

Seleccione una muestra de tamaño 10 y responda lo siguiente:

2. *¿Qué piensa acerca del valor de la proporción obtenida con base en esa muestra? ¿Cree realmente que corresponde al que puede tener la población? Explique porque si o porque no.*

3. *Si consideró que la muestra obtenida no representa a la población correctamente, explique qué podría hacer para obtener una muestra significativa.*

4. Comparar resultados de la muestra obtenida por cada uno con la de otros compañeros. *Anote los resultados de las proporciones obtenidas con las muestras seleccionadas (las suyas y la de sus compañeros) y explique cuál resultado podría estimar mejor el valor de la proporción poblacional que se desconoce.*

5. *Con base a lo que respondió anteriormente. (i) ¿Qué valores de los estimados son más probables que se presenten en la población? Explique. (ii) ¿En qué elementos de los estudiados (medidas de centralidad, intervalos de dispersión, tablas de proporciones, gráficas de barras) se basó para responder esta pregunta?*

(i)

(ii)

Parte 2: Inferencia de proporciones con una muestra de tamaño 30

Seleccione una muestra de tamaño 30 y responda lo siguiente

1. *Con base en la experiencia de lo que vio en la primera parte de esta actividad al tomar varias muestras de tamaño 30, considere la hipótesis acerca de la proporción o porcentaje de mujeres de la localidad de Kennedy que son acosadas en la calle que enunció en la primera actividad. Si cree necesario que la debe cambiar a otro valor, diga*

a que otro valor. Explique su decisión respecto a conservar o no la hipótesis primeramente enunciada.

- 2.** *Organice los resultados de las proporciones obtenidas con las muestras seleccionadas (la suya y la de sus compañeros). Explique cuál resultado cree que está estimando mejor el valor de la verdadera proporción que se desconoce. Para ello utilice información que aportan aspectos tales como las medidas de centralidad, intervalos de dispersión, tablas de proporciones, gráficas de barras, así como sus valores correspondientes.*

- 3.** *Explique si cambia de parecer respecto a la validez de la hipótesis enunciada al comienzo.*

3. Guía dos

Nombres y apellidos:

Curso:

Parte 3: Inferencia de proporciones con muestras de tamaño 100 o superior

1. *Plantee una hipótesis sobre la proporción poblacional de las mujeres que son acosadas, con base en una muestra de tamaño 100 e indique el valor de la proporción de la muestra seleccionada.*

Hipótesis:

2. *Comparta el resultado de la proporción muestral obtenida con la de sus compañeros observe el gráfico resultante. ¿Cambiaría su hipótesis inicial? ¿Por qué?*

3. *Con base en las proporciones obtenidas por todo el grupo ¿Cuáles consideran más probables y cuáles menos probables que se presenten en la población? Justifique sus respuestas teniendo en cuenta los elementos que aporta el estudio.*

4. *Concluya una relación entre el tamaño de la muestra y las medidas de dispersión e intervalos de variación obtenidos en las tres actividades.*

Muestra – medidas de dispersión

Muestra - intervalos de variación:

5. *Seleccione la proporción muestral que mejor pueda representar a la proporción poblacional y valide (argumentar si se rechaza o no) la última hipótesis planteada. ¿Entre qué valores podría estar la proporción poblacional de mujeres que consideran que ha habido casos de acoso callejero hacia las mujeres en la localidad de Kennedy con base en el estudio realizado?*

Juicio de valor de la hipótesis estadística (rechaza o no y por qué):

Inferencia sobre la proporción de mujeres acosadas en la localidad de Kennedy:

Inferencia de diferencia de proporciones

1. *Considere una muestra de tamaño 30 de mujeres que residen en la UPZ Patio Bonito y las muestras de tamaño relativa en las otras UPZ y halle las proporciones de las encuestadas que están de acuerdo con que en la zona ha habido casos de acoso. Realice una hipótesis de diferencia de proporciones entre (i) Patio Bonito- Américas, (ii) Kennedy Central-Patio Bonito. Ejemplos: la diferencia podría ser mayor que 0, puede no haber diferencia entre las proporciones, la diferencia de proporciones sería de -0,05.*

UPZ	Proporciones
Américas	
Kennedy central	
Patio bonito	

Hipótesis (i):

Hipótesis (ii):

2. *Suponga que para el caso de las diferencias proporcionales entre Kennedy Central y Patio Bonito se ha obtenido una diferencia negativa ¿Qué significa tener una diferencia menor que cero?*

3. *Considere una muestra de tamaño 100 para la UPZ Patio Bonito y las muestras relativas de las demás UPZ y halle las diferencias de proporciones solicitadas en el ítem 1. ¿Cambiaría algunas de las hipótesis iniciales? Sí/ No ¿Por qué?*

4. *Con base en los resultados obtenidos decida si acepta o rechaza alguna de las hipótesis planteadas sobre las diferencias de las proporciones. Justifique su respuesta.*

Juicio sobre la hipótesis 1: Patio Bonito - Américas

Juicio sobre la hipótesis 2: Kennedy Central – Patio Bonito

5. Suponga que el intervalo entre diferencias de proporciones entre dos UPZ es de $(-0.1, 0.1)$. Explique qué se puede concluir sobre la diferencia de proporciones en la población entre las dos UPZ para este caso.

4. Guía tres

Nombres y apellidos:

Curso:

Inferencia sobre las medias poblacionales con una muestra de tamaño 100

Considere como hipótesis la siguiente afirmación:

Las mujeres que habitan en la UPZ Castilla que dicen que en la zona se han presentado casos de acosos hacia las mujeres, tienen una edad promedio mayor o igual que 42 años.

1. *Tome una muestra de tamaño 100, calcule la media aritmética muestral (o promedio) de las edades de las mujeres encuestadas que dicen que las mujeres han sido víctimas de acoso en la zona UPZ Castilla, y reporte este resultado tanto en esta guía, así como a los encargados de la actividad.*

Edad media de la encuestada (años)

2. *Reporte las medidas de dispersión y los intervalos de variación que se obtienen en el aplicativo, que contiene los resultados de promedios de otras muestras de tamaño 100 de sus compañeros.*

Rango de edades	Desviación estándar	Media – desviación estándar	Media + desviación estándar	Media –2 desviaciones estándar	Media –2 desviaciones estándar

3. *Argumente si rechaza o no la hipótesis planteada y comente en que elementos estudiados (frecuencias absolutas o porcentuales de las gráficas, medidas de dispersión, medidas de centralidad, intervalos de variación, etc.) se basó para dar esta validación.*

Decisión sobre la hipótesis inicial:

Elementos tenidos en cuenta para la validación:

4. *Realice una conclusión para el parámetro estudiado, las edades promedio de las mujeres encuestadas que residen en la UPZ Castilla que están de acuerdo con el hecho de que en la zona se han presentado escenas de acoso hacia las mujeres, con base en los resultados obtenidos en el estudio de la muestra. Utilice la información suministrada por los intervalos de variación y por las gráficas ilustradas para justificar dicha conclusión.*

Conclusión

5. *Con base en los resultados obtenidos discuta sobre:*
- (i) *¿De qué edad(es) promedio sería menos probable encontrar una mujer que esté de acuerdo con que en el sector las mujeres son víctimas de acoso, y*
 - (ii) *¿Qué porcentaje de estas mujeres representan?*
- Para cada caso, mencione que tuvo en cuenta para dar su respuesta.*

(i)

(ii)

6. Considere lo aprendido en las guías trabajadas y responda ¿Cómo se puede mejorar la calidad del estimador, considerando el tamaño de la muestra?

7. Suponga que una mujer que va de compras cerca al barrio Portal de las Américas, el cual está ubicado en la UPZ # 46: Castilla. Si las distribuciones de las edades de las encuestadas y de las proporciones en la localidad de Kennedy se comportan similares a los de la muestra de tamaño 100 estudiadas (son significativas). ¿Cuál podría ser la probabilidad de que la mujer sea acosada? ¿Cuál es el rango de edad más probable de esta persona?

Probabilidad de ser acosada:

Rango de la mujer: